Factor Analysis

Under the Editorship of GARDNER MURPHY

Factor Analysis

An Introduction and Manual for the Psychologist and Social Scientist

By

RAYMOND B. CATTELL

RESEARCH PROFESSOR OF PSYCHOLOGY
UNIVERSITY OF ILLINOIS

HARPER & BROTHERS, Publishers
New York

FACTOR ANALYSIS: An Introduction and Manual for the Psychologist and Socialist Scientist.

Printed in the United States of America.

 B-B

Library of Congress catalog card number: 51-11893

To the Pioneers:

Charles Spearman *and* Louis L. Thurstone

I have not chanted verse like Homer's, no—
Nor swept string like Terpander, no—nor carved
And painted men like Phidias and his friend:
I am not great as they are, point by point:
But I have entered into sympathy
With these four, running these into one soul,
Who, separate, ignored each other's arts.

—Cleon

—R. Browning

CONTENTS

PREFACE

There have been moments when the experimentalist has looked upon statistical devices, such as factor analysis or analyses of variance, as fashionable appendages to scientific method, not likely to affect profoundly its methods or concepts. But it is now apparent that these contributions of the last half-century have come to stay and indeed to modify considerably the manner of thinking in which students are trained and the designs by which research is advanced. The simple, controlled experiment *may* continue to be the chief instrument of scientific method in the physical sciences. But the social and biological sciences, as soon as they recover from the initial slavish imitation of their older brothers, are likely to build a substantial part of their methods and concepts around these refined and powerful inventions in statistical analysis, which made their appearance simultaneously with the birth of inherent vitality in the new sciences.

Reasons for the appropriateness of this development in relation to the circumstances of the human sciences are discussed in Chapter 1. The development of factor analysis, as is well known, began in psychology, but spread quickly to education, sociology, economics, biology, medicine, and, recently, political science and physical and cultural anthropology. It has even rebounded into the physical sciences, notably into meteorology and electronics. However, at present the great bulk of accomplished work with this method lies in psychology. Yet it is in the neighboring social and biological sciences that more *urgent* demand for the method now exists; for comparatively great advances in organization are likely to reward the first applications of the method to an unstructured field.

In psychology itself factor analysis began in the structuring of personality manifestations, notably the abilities, and has moved lately into social psychology, as well as into learning theory, perception and comparative, dynamic, and physiological psychology. The fact that there is now no corner of psychology in which a student can expect to gain professional stature without at least a general understanding of factor

analysis may in some ways be a matter for regret. Individuals who can command both the artistic, intuitive, empathic skills of the good clinical or educational practitioner and the mathematical ability of the statistician are comparatively rare; yet factor analysis, because it is our most powerful wholistic method, is needed precisely in those areas which are handled largely by intuition, to carry it to surer levels. Thus clinical psychology, particularly, may gain greatly in the near future by the application of factor analysis in the form of P-technique, as described in the following chapters.

By this addition of rather complex techniques where clinical understanding previously proceeded on concepts and abilities of a primitive kind, applied psychology has merely repeated the history of applied physiology or medicine, where immediate clinical observation has given way to a new artistry with laboratory techniques. In psychology, though personal intuition and sensitivity are as valuable as ever, a mathematical sense of probabilities and degrees of interaction and evidence of functional unity has also become essential. The primitive discursive phase of psychology, in which the study became notorious as an academic refuge for students unable to face the mathematics of the physical sciences, is gone—though apparently some students have not yet been told. If the comparatively coarse problems of the engineer can be mastered only by a thorough grounding in mathematics, how much more exquisite must be the mathematical sense of the practitioners concerned with the prediction or control of human behavior?

Fortunately the degree of understanding of basic statistics and of such concepts as analysis of variance, multiple correlation, and factor analysis required of students or psychological practitioners properly to understand general issues in psychology is perhaps not on the same exacting level as that required of the research worker. Yet we have so far left the general student and practitioner with no alternative between struggling through the most thorough, advanced technical presentations of factor analysis, as set out for the advanced researcher, or remaining, on the other hand, in complete ignorance. A middle way, suitable for the majority, has never been provided. The researcher realizes a profound debt of gratitude to the three or four excellent, advanced, elegant, and exhaustive treatises on factor analysis to which frequent reference is made in this text, but it is unfortunately true that nine out of ten undergraduate students find them very difficult and are prone to conclude that the more complex correlation concepts are quite

beyond their powers. They acquire a severe phobia which is prone to issue in rationalizations doomed to stand inveterately between them and any real insight into the meaning of psychological measurement.

What looks like a formidable *pons asinorum*, apparently rooted in the discrepancies between the complexity of the subject and the level of the student's mathematical abilities, turns out, however, to be due to nothing except the lack of suitable immediate teaching devices. I have convinced myself by ten years of teaching factor analysis in courses on personality and social psychology, that a general understanding can be conveyed, by spatial presentation, in a digression of only two or three class periods, while three or four more classes will suffice to give the student confidence in practical working of the simpler factor analytic problems.

The only objection to this teaching solution is that one does not *have* in every course on personality, group psychology, economics, etc., time to give from two to six class periods exclusively to the subject of factor analysis. Yet, at present, one cannot avoid this digression as one does other necessary digressions, e.g., into physiology, anthropology, or genetics, namely by setting the student an assignment with a special synoptic auxiliary textbook; for only the above formidable textbooks exist. This book was also written to provide such an auxiliary text to other courses in social science in which a knowledge of factor analysis is necessary.

Actually, as the needs of the field were examined, the present book shaped itself to meet some three major requirements as follow. First, it sets out to meet the need of the general student in science to gain some idea of what factor analysis is about and to understand how it integrates with scientific methods and concepts generally. The first section of the book, comprising eight short chapters, has this purpose and is intended to be set as a reading assignment in courses primarily concerned with subject matter in which factor analysis is an incidental but necessary idea.

Second, it is intended as a textbook for statistics courses which deal with factor analysis for the first time, either as an appreciable part or as the whole of the semester course. For this purpose the second section of the book is intended to be added to the first: it picks up the subjects with which the student has gained an initial general familiarity in the first part and carries them to a higher level of precision and theoretical understanding. I think no apology is necessary for

designing a regular statistics course in a psychological rather than a logical sequence, i.e., for first whetting the student's appetite by a general acquaintance with the purposes and possibilities of factor analysis before proceeding to a more closely knit mathematical presentation. All experience shows that the student who is going to use statistics as a tool rather than a life study in itself is better taught by respecting his scientific interests more than the interests of mathematical sequence.

However, although the first and second sections combined give as comprehensive a study of factor analysis as can be gained in a first course on the subject they do not deal with particular topics as exhaustively as, say, Burt's *Factors of the Mind,* or with the whole field as comprehensively and mathematically as Holzinger and Harman's *Factor Analysis,* Thomson's *Factorial Analysis of Human Ability,* or Thurstone's *Multiple Factor Analysis.* The presentation does not concern itself with proofs of the formulas used; it avoids unnecessary adherence to technical mathematical modes of presentation, and it follows the sequence natural to the student's inquiring mind rather than the sequence of mathematical derivation and dependence. It is meant, in short, to make a broader highway to these excellent texts rather than to substitute for them. Experience indicates that the rarer students who will proceed to more advanced systematic, mathematical treatments in a second semeser course will approach later specialization with more zest through having his interests catered for in this way in the first semester. Easy stages and repeated contacts with the less simple ideas from several angles have been arranged, while the student has been freely directed to fuller treatments of statistical issues in the advanced texts at the points where discussion has to be cut short in the present introduction.

The third objective of this work is to supply a handbook for the research worker, the student, and the statistical clerk which will be a practical guide with respect to carrying out the processes most frequently in use. Factor analysis requires the skills of an art as well as statistical principles and it has been the author's objective to put at the reader's disposal some distillations of his own experience in a dozen years of factor analytic work in personality, social psychology, and more varied fields.

Some of the matters of interest to the research worker are dealt with in the second section—notably the art of rotation—but most

appear in the third section, in which the level of difficulty surpasses the two previous sections and in which the fullest complexities of the subject are frankly faced. Here the researcher will find some issues argued out that are not dealt with, or, in the writer's opinion, are not dealt with satisfactorily, in the classical textbooks referred to above. Even so the advanced theoretical level is maintained in the general area of scientific method rather than in regard to specialized mathematical and statistical knowledge, so that the objective of being readable and relevant to the science student as such is retained.

The instructor in a statistics course will not need to be told that it is essential for the student's grasp of subsequent chapters that he *learn by doing* the examples in the earlier chapters. But the student reading on his own is strongly urged systematically to work the graded examples provided and to answer, preferably in writing, the questions on concepts and theoretical issues.

To use the book as a laboratory handbook in factor analytic exercises or research it is necessary either to know it well enough to open at the right page or to make proper use of the handbook. In the interests of proper teaching sequence (including the need to ensure repetition of contact with complex ideas) the descriptions of actual computing processes are not arranged in a single condensed section of the book, and it would have been wasteful to gather them together again in some one section. Instead it seemed better to take care to arrange the subject index so that it could be used as a quick guide to the places where the working steps for particular processes are set out, and it is hoped that the practical worker will avail himself of this.

The writer wishes to record his indebtedness to Professors A. L. Comrey, J. Cohen, G. R. Grice, and especially L. L. McQuitty for reading the manuscript and making valuable suggestions on various teaching points. Miss M. Brannon and Dr. D. R. Saunders are to be thanked for drafting examples and checking calculations. The writer is especially grateful to the latter for several critical theoretical debates and to the former for exactness in preparing the manuscript and the proofs. Finally he wishes to thank Mrs. E. Morrison and Mrs. M. Henss, computing assistants, for help in preparation, and Miss D. Flint for her care in reading the proofs.

R. B. Cattell

University of Illinois
December, 1951

Part I

BASIC CONCEPTS IN FACTOR ANALYSIS

CHAPTER 1

The Place of Factor Analysis in Scientific Method

Scientific method, broadly viewed, wields two implements—that of experiment and that of statistical analysis. In its extreme form the first is characterized by its intention to control and manipulate in order to see how nature works. It wrenches a piece of nature from its setting and takes it into the laboratory to observe, usually without need of statistical aids. On the other hand, the extreme of statistical method proceeds without this interference and control. It observes events as they occur in their natural setting and, by analysis of relations in the measurements, attempts to find out what could otherwise be found by manipulation and control. Most actual researches employ some compromise and combination of these extremes. Nowadays it is usual to stress their unity, indicating, as Fisher does, that they merely differ statistically along a continuum concerned with the amount of error variance. But they differ also in situational setting and it is for many reasons important to recognize the scientific intention in each when in its pure form.

CONTROLLED EXPERIMENT

If we look at experiment more closely, we see that its aim is to establish causal or other relationships by the device of holding all conditions constant except the independent variable. By observing changes in dependent variables with controlled changes in the independent variable, the experimenter hopes to arrive at a law concerning their relationship. By the statistical approach, on the other hand, the researcher agrees to let many things vary at once, and aims by statistical analysis to isolate the particular relationship in which he is inter-

ested from those irrelevant relations which he chooses to regard, as far as his immediate purposes are concerned, as merely so much error.

The student familiar with the biological sciences will realize that in modern developments of scientific method some use of statistical method along with experimental method has become the rule. Indeed some statisticians think of statistical method as being coextensive with scientific method, being applied in some cases to natural and in others to artificial (experimental) observations. But actually science advanced for centuries on a basis of simple experiment, using perhaps elementary statistical logic but no statistical formulas; and it is only as more complex issues have been dared that statistical method has become essential. Even today in the physical sciences a pure experiment, with one design and one observation, may still be sufficient to test a hypothesis. But while most physical science research has become a combination of experiment and statistics, the developments of the social sciences (and some aspects of meteorology and astronomy) have reached a point where the testing of a hypothesis frequently presents as pure an instance of statistical, mathematical analysis of unmanipulated situations as the early physical sciences presented of simple experiment.

An illustration of the adaptation of method to varying degrees of control of influences can be perceived by comparing two investigations —one into the physical relation of the length of a pendulum to its time of swing and one into the relation of vocabulary to length of schooling. In the former the researcher will vary the length of pendulum, while keeping everything else constant, *as far as he knows,* and plot the relation of length to period of swing, as in Fig. 1. The italicized phrase is important, for advocates of the superiority of experimental method are apt to forget that the researcher may not *know* all the things to be held constant. For example, in this instance he might hold constant the weight of the bob, the temperature, amplitude of swing, shape of bob, degree of exhaustion of atmosphere, but not know that it is important also to make all observations at the same altitude.

STATISTICAL ANALYSIS

In seeking similarly a scientific law relating length of schooling to size of vocabulary, the investigator is unlikely to use the experimental method—in the sense of controlled experimental method. For he can-

not compel some individuals to stay at school longer than others while he tests their vocabulary and he cannot control the rest of their lives so that, except for schooling differences, they are exactly the same. Furthermore, he even has one influence—age—which is inextricably entangled with one of his variables (which he may wish to call the independent variable), for those who have been at school longer will in general be older. This difficulty can scarcely be avoided, since it will hardly suffice to start some at school when they are younger, because the effect of schooling will depend upon the age at which it occurs.

The researcher is therefore likely in this case to attempt no manipulation but to take cases with the natural variability of schooling and vocabulary which he finds given in nature and to apply statistical,

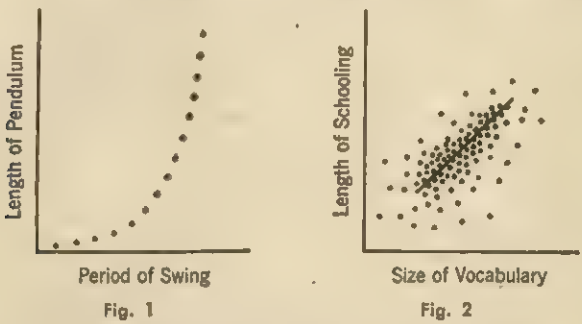


DIAGRAM 1. Experimental Observations with and Without Irrelevant Influences.

analytical methods thereto. In the simplest approach he is likely to plot a graph of the relations just as he did in the pendulum experiment, except that he *finds* his cases instead of making them. The two statements of data will be as in Figs. 1 and 2 of Diagram 1.

In the second case, individuals with exactly the same length of schooling show some scatter in vocabulary because the investigator has failed to control other conditions, such as health, intelligence, and interest in reading, which also partly determine vocabulary. He can obtain the best *estimate* of the relation of duration of schooling to vocabulary by taking so many cases that he can assume that the groups on which he bases the average points are equalized, by random sampling, with respect to these other conditions. This is essentially the

relation that he would get if he put the best fitting line[1] through the swarm of points, as in Fig. 2, Diagram 1. As students familiar with correlation will recognize, the correlation coefficient is an attempt to express the degree of confidence with which a recognized linear relation can be used in predicting from one variable to the other.

It is probably unnecessary, for the reader having any acquaintance with scientific work, to digress into explanations of the theory or practice of reducing the effects of chance error by taking a large number of observations. However, it may be pointed out that systematic and chance errors arise broadly from two sources: sampling error, i.e., the individuals tested not being truly representative of the population; and error of measurement due to some influences in the experimenter, the subject, or the test intruding upon the thing that the experimenter sets out to measure. Just as dirt has been defined as matter in the wrong place, so errors are only real influences (and often very interesting influences) operating in the wrong place. The trembling hand on the ruler is error to the engineer but significant data to the doctor. Random error of measurement covers, therefore, all the influences one is *not* interested in measuring at the time. By taking many observations *systematically* chosen with respect to what one *wants* to measure and *randomly* with regard to all else, one can generally, but not always, obtain a statistical mean which does not differ *systematically* from the true value.

It would be too large a digression from our pursuit of factor analysis to make any thorough examination of the advantages and disadvantages of putting greater emphasis respectively on experimental and statistical methods—in situations where choice of emphasis is *possible* in research design. Yet some comment must be made since undoubtedly psychological research and the whole development of psychology have suffered considerably through failure to realize the relative utilities and mode of interaction of these methods. Unquestionably, the experi-

[1] The line fitted by least squares as a best fit may be regarded as lying midway between the two regression lines which converge upon it as correlation becomes perfect. One of the regression lines has the slope obtained by taking mean sizes of vocabulary for different lengths of schooling. If the variations in schooling and vocabulary are first expressed in standard scores, the slope of this regression line, i.e., the *tangent* of the angle which it forms with the horizontal line, actually equals the correlation coefficient. The best fitting least squares line shown in Fig. 2 above, and which would normally be used for getting reversible prediction, i.e., best equivalent values in the two variables, is a line dividing the angle between the two regression lines corresponding to the correlation coefficient.

mental method is more immediately attractive, notably in the positiveness of the results which it yields. Thereby the scientist ideally obtains a so-called causal law—that when A happens, B will undoubtedly follow—whereas by statistics he obtains only a law of probability, an estimate of the likelihood of B happening if A occurs, as illustrated by the scatter of points in Fig. 2 indicating such and such a vocabulary for a given value of schooling duration. The connection of a given change in the dependent variable with a given change in the independent variable is absolute, because there is no fuzziness about the line in Fig. 1. The fuzziness in Fig. 2 means that a number of extraneous influences, the characters of which the investigator generally does not know, are entering into the determination of the change in the dependent variable. Indeed in these circumstances the investigator cannot even be sure, unless time sequence is involved, which *is* the independent and which the dependent variable.

Partly because of this greater feeling of certainty and control and partly, one regrets to say, because some psychologists have no better conception of scientific method than slavishly to imitate elementary physics, experimental methods have been mechanically and unsuccessfully applied in areas of psychology and the sociobiological sciences where statistical approaches would have been more appropriate and effective. As Brunswik (**8**)[2] points out in his analysis of methods and their related concepts, most of the laws of psychology are likely to remain probability laws; and in the last resort in science, as Lecomte du Noüy has reminded us (**99**), *all* scientific laws are statements of probability rather than of infallible sequence or causation. We do not know enough about all the influences that may affect a particular situation—so vast is the universe—to make a prediction of an infallible nature from our own narrow range of experience of invariability. Someday the falling stone may start spontaneously to move upward. As pointed out above, the experimenter can never be certain that he knows all the influences that he is or is not holding constant.

Although it is thus a delusion to think that experimental method is superior to, or different in kind rather than in degree from, the statistical method, there are certain advantages of outcome and conveniences of procedure peculiar to each method. If it *is* possible to control the situation by masterly interference, and if one already has such familiarity with the phenomena that one knows what influences

[2] The boldface numbers in parentheses refer to the Bibliography on page 443.

are likely to need control, the experimental approach saves much labor. In psychology, this is possible with such restricted fields of interest as the special senses or perception, but in the present state of social psychology and personality study, the assumption would be a quite unrealistic basis for research. For example, it rarely suffices in social science to experiment with just a single individual, since we are practically never in the situation of knowing what factors in individual differences might or might not affect the phenomenon being observed.

The problems in psychology that are of the greatest practical importance today and—in the present writer's opinion—also of the greatest theoretical beauty, are those concerned with the behavior of the total organism i.e., personality study, or with the interrelation of individuals and organized groups i.e., social psychology or sociology. The vital issues in these fields cannot easily be brought alive into the laboratory and it is not surprising that the purely experimental approach has barely touched the surface of them. Indeed, experiment in the narrow and strict sense of, say, the Society of Experimental Psychologists, has yielded good results only in a small corner of psychology. Phenomena such as a schizophrenic breakdown, the rise of a new political party, or the genesis of a thunderstorm cannot be studied by the pure experimental method. Both the number and the nature of the influences at work puts them beyond control. It is in this situation that statistical, wholistic[3] methods come into their own.

With the dawning recognition that the relative emphasis on experimental and statistical methods has to be very different in the biological and social sciences from that in some of the older sciences, and that the newer sciences have to invent their own special submethods and unique combinations of methods, there has developed a vigorous application of statistics to the above mentioned areas of naturalistic observation *in situ*. For it is incorrect to suppose, as experimentalists have sometimes done, that the only possible alternative to their own rigorous but narrow methods is the loose farrago of unverified observation and bottomless speculation found, for example, in much clinical and cultural anthropological research. Today the methods of unaided personal clinical observation and cultural anthropological description have reached the limits of their powers to bring substantial advance.

[3] As will be seen later, factor analysis is essentially a wholistic method in that it constructs statistically from a host of variables (observations) the important wholes which need to be taken into account when seeking laws of interaction.

But at this point the more ingenious application of statistical techniques has begun to bring as much acceleration in personality study and social psychology as the addition of the microscope to the unaided eye brought in the history of biology.

FACTOR ANALYSIS IN STATISTICAL METHOD

Having dealt with the relation of experimental method to general statistical method we can now turn to the special role *within* statistical method of *factor analysis*. Elementary statistical method is mainly concerned with finding the means and standard deviations (scatter) of measurements or with discovering whether the differences between various means and sigmas are significantly beyond chance. It is directed to separating the variability associated with uncontrolled variables from the variance due to the influences in which one is interested. Such statistical approaches, whether they operate in experiments that are actually controlled, as when we study plant growth with two different kinds of applied fertilizer, or whether they operate on cases that have merely been selected from nature's experiments, as when we compare physiological measurements of schizophrenic and nonschizophrenic individuals, continue to have in common what may be called arbitrary choice of variables in relation to a single dependent variable or concept.

By arbitrary choice of variables we mean that the experimenter selects on the basis of his own hunches the controlled or uncontrolled variables that will best test the hypothesis he has in mind. For example, in the above experiment on schizophrenia he may select schizophrenics by the variable of being or not being diagnosed as such by a group of four psychiatrists or by falling below a certain ratio of "adaptive" to "verbal" intelligence test performance, or by talking to imaginary voices and so on. And he might choose red blood count as a physiological variable because his hypothesis is concerned with the notion of metabolism rate in the brain. An economist would similarly choose arbitrarily, i.e., on personal reasoning, such a variable as money interest rate or bankruptcies per year as representatives of the state of the trade cycle in some study relating other influences to the trade cycle. These choices might be wrong, i.e., the variables might be quite poor or misleading indicators of the condition, concept, or whole they are supposed to represent.

The branch of statistics with which this book is concerned, namely,

factor analysis, is a more radical departure from the statistics associated with the experimental tradition, in that it does not accept arbitrary choices as to what are the important variables in any field. Nor is it satisfied, as is analysis of variance,[4] simply to answer yes or

[4] Analysis of variance has been taken as the typical and fullest expression of that major branch of statistics which represents the chief technical development at the present time *outside* factor analysis. Therein, the student may be reminded, several influences are brought to bear upon a single variable. For example, the mean and sigma in respect to reaction time might be determined for groups which differ in length of training, body build, and amount of alcohol taken before the experiment. These influences (independent variables), each broken down into three or four grades, can be applied in cross classification (e.g., Latin square) row and column form so that each possible combination (or a fraction of the possible combinations) of high and low with regard to one is represented once with each high and low in another. Thus if there were only the first two variables above and each was present in two degrees, we should have four experimental groups—long training, light body build; long training, heavy body build; short training, light body build; short training, heavy body build. This necessity for obtaining all combinations in the ideal design makes analysis of variance more suitable for experimental than naturalistic, statistical method, for one cannot be sure in nature of obtaining samples of all possible combinations. It should be understood that in what follows we refer not to that analysis of variance which merely contrasts several influences ("effects") but to that in which each influence, broken down into several grades, is one side of a Latin square.

When the calculations are through, one has gained the information that the given dependent variable (reaction time) does or does not significantly respond to differences in some or all of the independent variables—training, body build, etc. The variance is broken down into within-group and between-group variance, which tells us whether the groups differ significantly in their means and whether or not we can entertain the null hypothesis that they are behaving as chance samples from a larger population. If they cannot, then the influences we have applied in the form of the independent variables can be said to have some significant effect upon or association with the dependent variable. Factor analysis differs from analysis of variance principally (1) in yielding evidence as to the *strength* (not the mere presence or absence) of association between two variables, (2) in requiring no suppositions as to which are dependent or independent variables; (indeed it yields evidence on every combination between the variables instead of between only the dependent variable and the various, distinct individual, independent variables), and (3) in revealing whether the independent variables as assumed in the analysis of variance are in fact (a) mutually independent and (b) the really important independent influences in the field. Two independent variables in analysis of variance might indeed be essentially the same variable in disguise, but the method would not reveal this. Factor analysis groups the numerous possible variables in the fewest possible single wholes or wholistic influences. Analysis of covariance, however, throws *some* light on (a).

This third difference may need further explanation. As factor analysis is usually applied, the independent, controlled, or criterion variable or variables are not factorized in with dependent variables. In that case the independence or mutual entanglement of the independent variables which it is proposed to use in analysis of variance are revealed by a *prior* factorization. Moreover it is rather

no to the question of whether a change on one variable is associated with a change in another. It goes further, both to determine the *degree* of the association and to pick out the essential wholes among the influences at work. For a statistically significant difference of means may yet constitute so slight a degree of association as to be psychologically insignificant.

Having demarcated factor analysis from the chief complementary statistical method (see Burt (**14a**) for an extended analysis of the possible relations of these methods) we can now appropriately develop more fully its specific role in scientific method. At this point, it is first necessary to define more exactly certain aspects of both science and factor analysis. The aim of scientific method is to discover facts and the relations among facts. Most facts, e.g., that ice melts at 32° F, turn out on closer logical analysis to be themselves relations among factual fundaments; and for practical purposes we can say that scientific method is concerned with isolating relationships, i.e., with bringing order by discovering predictive laws.

SCIENTIFIC METHODS

Philosophers of scientific methods (**1, 38, 50, 87**) have spoken of four kinds of order: (1) constant conjunction of properties, as used in recognizing a chemical element; (2) causal order, i.e., an invariable

---

uncommon in classical factor analysis to apply factorization to conditions of the environment instead of to attributes of organisms. For example, industrial psychology may factorize a dozen or more human performances, in some defined conditions of temperature, light, and humidity but rarely thinks of factorizing the latter. Nevertheless the conditions *could* be factorized (as they covary from day to day) and the legitimacy of entering them as independent variables in a later analysis of variance could thus be tested, with the discovery of the true independent influences to be used. In condition-response factorization (Chapter 20), however, the factorization of dependent and independent (condition) variables takes place in a single experiment.

Although some analysis of variance experiments can thus be handled at times by factor analysis, and should more frequently be preceded by factor analysis, the objectives of the two methods are in general so distinct that there is no rivalry in regard to their claims for employment. Analysis of variance, even when it deals with numerous conditions or independent variables, which happen to be of a kind that could be taken into a factor analysis, is interested in the outcome on a *single dependent variable*, whereas factor analysis is principally concerned when many responses or dependent variables are being measured. Analysis of variance, moreover, is uninterested in the sources of variance—the individual differences—in the organisms themselves. That is just so much error variance. But factor analysis is interested both in the variance in the organisms and in the environment, though classically it has generally been brought in only when the former is the object of study.

sequence, as that failure of oxygen supply causes death; (3) numerical relations; and (4) relations among constructs, not all the elements of which can be directly observed, i.e., order among theoretical entities, each an abstraction from observed data.

The methods by which science seeks to establish these relations have been set out by Bacon, Mill, and later students of method under some variety of the principles of agreement, of difference, of residues, and of concomitant variation. They issue in the experimental and statistical research designs we have already mentioned. All are concerned to establish covariation of variables, either without regard to special time sequence as in most numerical laws or constant conjunction of properties, or with regard to time sequence as in causal laws.[5]

The extent to which a hypothesis is deliberately invoked in the search for relations will vary. Some hypothesis is involved, at least implicitly, in choosing to observe certain possible fundaments for a relation rather than certain others. But the hypothesis, implicit or explicit, may vary from one which states a simple relation between observed variables to one which states a relation between complex entities of which the observed variables are only distant representatives. A continuum in regard to complexity of relations and extent of reasoning by analogy exists between the simplest empirical law, e.g., that bodies of different weight fall at the same rate, and the most elaborate hypotheses, e.g., that repression is a cause of neuroses. All laws and hypotheses are statements of expected relations, and the better ones are distinguished by being more precise as to what the verbal hypothesis *means* in terms of expected observations. It is a common weakness of hypotheses in the social sciences, incidentally, that they deal with elaborate or pretentious verbal concepts which cannot be related with certainty to the variables, the observations upon which are nevertheless supposed to test the theories. Such theories

[5] The reference to time sequence can perhaps be most readily illustrated by a simple physical example, as expressed in the formula covering Boyle's and Charles's laws, viz: $\frac{PV}{T}=C$. This expresses relations regardless of whether P, V, or T is first changed. On the other hand the statement that if I put a match to gunpowder, it will explode (or the statement of the irreversible chemical equation) is a generalization about an invariable time sequence. Of course, if a gas be compressed adiabatically, the above equation also modifies: there is then a dependent and independent variable and a law involving a certain time sequence.

may even direct the attention of the observer *away* from noticing more real connections between the variables.

The historical fact that hypotheses can delay the perception of lawful relations quite as readily as they can facilitate it tends to be overlooked because we are acquainted largely with the true hypotheses which survived. How much did the theory that malaria is due to marsh gas delay the perception of the role of the mosquito, or the chemical theory of phlogiston interrupt the discovery of oxygen and the law of indestructibility of matter? Often the heated partisanship between alternative theories A or B has long directed attention from the observation of palpable regularities which would show that hypothesis C is the true mechanism behind the scenes!

THE PROPER ROLE OF HYPOTHESES

Although it may seem unimportant to the novice as to whether we state the law we expect to find *before* observing covariations or whether we observe the events and then *seek a law to fit them*, the history of science shows that it is most important to have a nice compromise here. Classical accounts of scientific method, more concerned with intellectual pomp than historical and psychological truth or present research fruitfulness, have overstressed the importance of hypotheses. It *is* possible to observe covariation and to develop laws *without* theories. An observer may note the curious fact that the Nile brims its banks in the summer, that steam lifts a kettle lid, that people who catch malaria are often cured of general paralysis, or that schizophrenics practically never suffer from epilepsy. The observed covariation then leads to many possible theories. In a sense, as stated, it is true that one has a theory even in choosing facts to observe, but it is a theory of a very *broad and flexible kind*—and in historical instances has sometimes been little more than a profound sense of wonder and a disinterested curiosity. The best researchers are those who are not so passionately wrapped up in their theories that they fail to observe events on the side. And these unforeseen by-products of research generally stand up to further examination better, precisely because they are undistorted by any desire of the experimenter to produce them. The essential thing to observe is any and every evidence of *law*, i.e., of orderly covariation, in the field concerned.

Now this analysis of values in scientific research is important to a proper appreciation of factor analysis because the latter offers a com-

prehensive and sensitive method of expressing quantitative relation from the observation of *covariation.* While the method is more sensitive to quantification of relationships than analysis of variance, it is not sensitive enough to lead to *exact* or complex equations relating the dependent to the independent variable. *Classical* factor analysis is restricted to the study of covariation as it occurs naturally. But a development of factor analysis called condition-response factorization, described in Chapter 20, permits its use with *controlled* covariation. Factor analysis provides also a method far more free than most methods from the necessity to elaborate rigid hypotheses. It is the ideal method of *open exploration* in regions unstructured by present knowledge. In embarking upon a factor analysis one need have no more definite idea than Columbus had of America in regard to what may be found. It is sufficient to hypothesize that *some* structure lies there. Columbus interested his backers by telling them he was sailing to China, and some similar explanation in terms of familiar concepts may be necessary for our modern research foundations; but it is questionable whether these concepts are help or hindrance in factor analytic research itself.[6]

FACTOR ANALYTIC PROCEDURE

To see this more clearly it is necessary here and now to glance at the actual procedure in factor analysis, though the full understanding of the steps in this procedure may only become apparent by later reading in this book. To begin with, we take measures on a number of variables in a certain field, e.g., a set of forty diverse personality tests measured on five hundred people, or two or three dozen indexes of trade activity measured every week for two or three years, and work out all possible correlation coefficients among the variables (measures) to see to what extent they covary. For example, in the personality variables we might find that sociability measures covary positively with health measures but that intelligence measures are unrelated to either of these, i.e., the more intelligent person is not necessarily more healthy or more sociable. Factor analysis, carried out on the correlation coefficients, shows us how some variables can

[6] The same is true of any truly exploratory or creative research methods. It is questionable whether Freud would have gotten any support in 1900 for a proposed research on the Oedipus complex or whether Franklin, at the time of his electrical experiments, would have received substantial help from power, light, and haulage companies.

be grouped together because they behave in the same way, and it proceeds to delineate new independent, underlying factors which may be responsible for these groupings. For example, it might pick out a group of variables all involving quickness, activity, and sensitiveness whose common variation could be traced to differences in activity of the thyroid gland. The latter is a factor behind the actually measured performances, and in general we find a relatively small number of such major factors responsible for a substantial part of the variation in a relatively large number of variables. Factor analysis might therefore almost as well be called factor synthesis or at least variable synthesis, for although it analyzes out the distinct factors at work among the variables, it also groups the variables together in ways which permit one to synthesize new entities. These new entities are now themselves to be considered as variables—though far fewer than the initial raw variables—which can be used as hypothetical causes, intervening constructs, or independent influences behind the more numerous and bewildering mass of raw variables.

In his recent review of the development of psychology, Klüver (**82**) notes that factor analysis has opened up new fields and concepts, but he conjectures as to how this can be since Thurstone, one of the leaders in this field, calls attention to the fact that "factor analysis has its principal usefulness at the borderline of science" where fundamental concepts are still lacking and crucial experiments cannot be easily devised. This is a very sound appraisal, providing we substitute base for borderline, but it overlooks possibilities of application beyond the classical use of factor analysis, possibilities which create the apparent paradox that factor analysis belongs both to the very earliest stages of research and the very last.

It belongs to the earliest stage of research because there is no point in working out—or rather hoping to work out—precise laws about the relations between variables until we have chosen the *significant* variables, i.e., the important, major influences, between which regular relations are likely to exist. The factor analyst is suspicious of choosing the important variables a priori, no matter how self-evident their significance may seem to the experimenter. He would like to find the real independent factors, the true functional unities, i.e., the independently acting influences, before entering an experiment with them. He believes that to find relations between some variable, e.g., the maze-learning speed of a rat, and some factor in the situation,

e.g., the strength of the rat's hunger drives, is of little value if the measures of the factors—hunger and goodness of learning—are some unrepresentative, contaminated, and logically uncertain measures, e.g., a measurement of hunger drive in terms of hours since eating (See Chapter 20). Again, the experimenter who chooses his variables on mere hunches without prior factor analysis may find that in his blindness he has taken two or more variables which are really different manifestations of the same thing. It is necessary first to *find* what relatively independent functional unities are operative in the situation and *then* do experiments on them. Each true functional unity can then be represented in the experiment by taking a variable or more commonly a combination of variables which are shown by the factor loadings to provide the cleanest measure of that factor. Of course the factors will not be *entirely* independent, else there would be no point in doing experiments on their relations to find laws connecting them. But they will be functionally independent, as say, temperature and pressure are in physics, or time and effort in learning. Controlled experiment following preliminary structuring of the field by factor analysis is thus in a different methodological category from controlled experiment with mere a priori or random variables.

Particularly in the biological and social sciences the researcher is presented with so bewildering a multitude of possible variables that unless he first factorizes to find the inherent organization or *structure*, i.e., to find which surface variables are representatives of more significant, less numerous underlying variables, an immense waste of effort could (and does!) take place. In these fields we are like the crew of a ship approaching some strange coast through a fog. It is easy to seize on some arbitrary, transient point of visibility and still easier to convince ourselves that it proves the existence of structures created by our own imaginations on the basis of pretentious hypotheses. Factor analysis, however, comes to our rescue as a kind of radar to avoid both the trivial and the unreal, for it gives us—however roughly at first—the shape of the real structures hidden in the swirling multiplicity of variables.

To illustrate this metaphor more particularly, we may glance at the situation in research on human personality. Innumerable investigators are concerned to find out how this, that, and the other environmental circumstance affects personality. But they do not know what to measure (in the scope of the few dozen variables to which

they must restrict themselves) effectively to *represent* the main aspects of personality, any one of which might be affected. Here the discovery of the primary abilities (**125**) and the primary personality factors (**22**) provides a real frame of reference—a set of measures of the most important functional unities about which to organize experimental findings. Or again within social psychology, the study of group dynamics in the broadest sense is concerned with discovering how different systems of leadership, communication, population selection, etc. affect morals, aggressiveness, effectiveness of group performance, and other dimensions of groups. Before findings of any permanence can be obtained, i.e., of the kind which can accumulate about definite, reproducible, universally meaningful aspects of groups, it is necessary for basic research to proceed upon the factorization of a wide variety of group performances and observations to determine the most important functional unities in the behavior of social groups.

The formation of hypotheses about these discovered unities, for further experiment, is on a totally new, and a higher, level of scientific importance compared to hypothesizing at the level of already familiar variables. In social phenomena, for example, the surface variables have been known to mankind so long that most relationships which intelligent men could perceive have long been perceived. Factorization brings forth a new order of variables and concepts on the relations among which it is richly rewarding to begin forming hypotheses—hypotheses which can be of wider reference and import than those built intuitively on surface variables.

So far only a passing reference has been made to condition-response factorization, since what is peculiar to factor analysis is most clearly shown in what may be called classical factorization—the accepted and widely used form of the method. But special developments are described later (Chapter 20) which permit factor analysis to be used along with experimental control of independent variables. This permits factor analysis to serve the same area of scientific work as is served by analysis of variance, while still retaining the greater structuring power of factor analysis.

Factor analytic investigation has its function predominantly, therefore, in *basic* research to provide the measurement foundations for later special problems in pure and applied research. But there is no intention pedantically to insist that all experimental research in a new field must begin with factor analysis. Due to circumstances it is de-

sirable on some occasions and necessary in others to carry out experiments in which we have to take isolated, a priori variables, without knowing how they fit into more basic factors.

FACTOR ANALYSIS AND OTHER STATISTICAL METHODS

Factor analysis is a wholistic method in that it aims to discover and deal with the more massive functional and organic wholes instead of losing research perspective in a mass of atomistically conceived variables; but it is not the only statistical method with this objective. Such statistical devices as multiple correlation, partial correlation, and the discriminant function, lying between the more passive statistics of experiment (comparison of means, analysis of variance) and the active exploration of organization by factor analysis also have this objective to a lesser extent. They deal with the grouping of variables in relation to some wholistic single effect and they also attempt to give some account of where *all* the variance in a particular phenomenon goes to or issues from. That is to say, they also seek for a more *complete* account of the influences at work.

Although these related methods will be discussed again in Part III, it is desirable for the sake of orientation to refresh the student's memory regarding their nature and purposes while contrasting them with factor analysis. In partial correlation it is our aim to eliminate the effect of one (or more) contributory influences to a correlation to see how much remains due to the influence which most interests us. For example, we may seek to discover the correlation of intelligence with general information when age is held constant. Factor analysis achieves the same end as partial correlation, for that factor analytic equation which we shall recognize later as the specification equation gives the correlation of a particular performance separately with *each* of a number of factors when the others are held constant. It differs from partial correlation in that this procedure holds whole *factors* constant where the former holds *variables* constant. Indeed it is one of the chief weaknesses of partial correlation that it sometimes blindly attempts to hold constant a variable which is intrinsically part of the same functional unity as the dependent variable.

In multiple correlations, knowing the correlation of each of several variables with a criterion as well as their correlations with one another, we attempt to obtain a weighted composite of the variables that will give the best possible prediction of the criterion. This also is

done, in a somewhat more fundamental way, by factor analysis. But again factor analysis is more systematic in that it first *groups* the variables, to give estimates of independent functional unities, and then predicts the criterion from these. In some particular applied problem, having variables which do not occur widely in other studies, it may be quicker to use multiple correlation of arbitrary variables. But since factors correspond to psychological traits, about which a good deal may be known from other researches, it is possible to predict with more insight and ultimate control if we first factorize. Thus in factor analysis generally it is clear that the experimenter is made to respect the organic relations inherent in the material. He cannot attempt to hold constant what is an essential part of the thing that is varying, nor can he confine in a hodgepodge of naturally unrelated or unduly overlapping variables, measured variables that are not part of the natural unity of a factor.

The discriminant function (**127**) is also a wholistic device in that it tells one how to combine (i.e., by what weights to add) a set of variables to give a total which will show the maximum difference or discriminating power between two groups, e.g., what total picture best distinguishes a sane from an insane individual, or society at a boom period from society at a depression. This differs from factor analysis again in that it is arbitrary in its combination of variables, at the very least to the extent that the experimenter *has to choose his two types first*, e.g., to decide which is a boom and which a typical depression. This applies equally to Rulon's "generalized discriminant function," using several "types."

This last is a weakness of the discriminant function method if one is searching for true organic wholes. For although the combination of weighted variables which will best distinguish one group from another *may* be, and probably sometimes is, the expression of a single factor, this cannot be assumed. For example, one might seek to define intelligence—indeed the approach has been tried—by noting the tests which best distinguish mental defectives from normals or geniuses. But the mental defective group is likely to differ from the normal by other factors than intelligence. For example, their grouping apart socially has added such characteristics as poorer physique and more antisocial tendencies and these would have to be given weight in any discriminant function giving a maximum separation of the two groups.

In general, the alternative statistical tools just considered differ

from factor analysis in arbitrarily fixating their attention on particular variables and a particular criterion. They are just as *effective* in giving a prediction of performance *within one restricted research*—indeed they involve less computation and avoid the error of estimate of intermediate variables or factors. But they do not contribute as factor analysis does to something beyond immediate prediction—namely to scientific understanding of what basic influences are operative. That is to say: they are scarcely tools of investigation. They are therefore used more widely in particular *applied* problems than in pure science, for they contribute little or nothing to prediction in terms of *scientific concepts*.

The above discussion has emphasized the role of factor analysis at the exploratory beginnings of science, in unearthing the functional unities for further study, but we must next recognize that it has a function at the level of highly finished research as well as at the basic exploratory level. It may indeed, *follow* such methods as testing the significance of difference of means, or the application of analysis of variance to a collection of means, instead of preceding them. For while we have argued that in certain designs it is profitable to apply analysis of variance only after preliminary factor analysis has shown what unitary influences are best chosen to enter in the analysis of variance design, it is also true that the latter method has a place as a scouting method *preceding* factor analysis. For factor analysis (and correlation generally) is concerned with *how much* relationship exists while the former merely indicates whether *any* significant relation exists. Factor analysis is a finer quantitative tool, giving answers appropriate to a more advanced stage of research. But it has a role at more advanced stages also because it permits a more complex and detailed hypothesis to be tested than is possible by other statistical methods. Analysis of variance shares the ability to handle several independent variables and complex interaction effects at once, but factor analysis can indicate both how many are in action and what the magnitude of their action is. Analysis of variance can yield evidence on interaction effects of one kind, but factor analysis yields more extensive evidence of other kinds of interaction, notably extent of assistance or opposition of influences in their effects. Consequently with factor analysis we can experiment with hypotheses that extend to statements about the *number* of factors at work in a situation, the

nature of the factors, their degree of *interaction,* and the *magnitude* of their influence.

On the other hand, and in the role of an exploratory method, factor analysis has the peculiarity, among scientific investigation tools (as indicated above and explained more fully in Chapters 8 and 18), that it can be profitably used with relatively little regard to prior formulation of a hypothesis. Like radar turned upon a fog—to continue our earlier metaphor—it necessarily reveals to us whatever organization or structure is present. This freedom alike from direction or distortion by hypothesis also exists, it is true, in other methods, though to a lesser extent. For example, we can look for a difference of means in some experiment, with or without a hypothesis as to what might cause such a difference. But the complex structure of statistical relationships revealed by factor analysis has decidedly more intrinsic meaning than a difference of means and is chosen from a far more vast array of alternative possibilities than the mere answers more or less.

Starting with measurements on two or three dozen variables, a factor analyst can thus, without hypothesis formation, arrive at the highly structured answer that there are, say, five factors at work, that their natures are such and such, that they are correlated among themselves in such and such a manner and have certain specific relative magnitudes in respect to their contribution to the variance of a particular variable or of most variables. Of course, it would generally be a richer contribution to understanding if he had first been able to formulate an exact hypothesis to this effect from contributory evidence gathered in other fields or by other methods. For example, if a hypothesis as to human abilities is set up from brain physiology observations and is then confirmed by factor analysis of ability performances, the result is more impressive and has wider connotations than if it appeared for the first time in a factor analysis. But the detailed hypothesis is not essential to the design of the experiment. Of course the factorist enters with *some* hypothesis even when he seems to enter with none. He enters an experiment with the hypothesis *that some structure exists to be discovered,* just as Columbus set sail with an hypothesis that some land existed to the west. Later, when we study the techniques governing choice of variables we shall see that the choice of variables may be to some

extent affected by the hypothesis, but there are also general rules for the choice of variables which have to be observed regardless of particular hypotheses or, perhaps one should say, in spite of them.

SUMMARY

To summarize; a scientific research method in general can be basically evaluated chiefly in regard to four characteristics. 1. Does it use experimental control or naturalistic observation? This concerns the continuum of interference and control discussed above. 2. Does it arbitrarily choose[7] a collection of variables whose relations are to be investigated or does it establish their importance as unitary factors by their empirical connections? 3. Does it require hypotheses or does it tend rather to produce hypotheses? 4. Does it force the experimenter to make prior assumptions as to which are dependent and independent variables or does it leave this causal relationship unassumed and open to later proof?

Factor analysis corresponds to the second answer in regard to each of the above questions—except that the stimulus-response factorization permits some controlled variables. Its strength lies in making fewer assumptions in 2, 3, and 4 and in requiring less physical control of the situation in regard to 1. It can test *or* produce hypotheses. In the biosocial sciences, where control is sometimes completely ruled out, where hypothetical assumptions are almost as frequently misleading as helpful, where the variables representing important structures have still to be found, and where the direction of causation is anyone's guess; this method, even with its present technical bugs, is the most useful tool we have.

The main value emphasis in the above discussions has deliberately been placed on factor analysis as a means of investigating nature, i.e., as a tool of basic research. But many people will become acquainted with it for a second reason: that it is an instrument of prediction in relation to specific problems, i.e., a tool of applied research. Prediction does not mean here prediction to be used as the test of an hypothesis—which it can also be—but prediction as a practical, routine procedure in applied science. Problems ranging from the estimation by vocational guidance tests of a person's probability of success in an occupation to prediction in economics of a boom or

[7] Assumptions about a variable representing some abstract theoretical entity are included in our term arbitrary choices.

the assessment in sociology or political science of the likelihood of a particular country being involved in a war can be handled by the specification equation (Chapter 6). As more knowledge of factors and their operation accumulates in relation to these issues, its predictive function will become more important, and its use for this purpose will need to be learned by an increasing number of practitioners; but at the moment its greatest importance is to the researcher and student concerned with that scientific understanding of which effective practical prediction is a mere by-product.

Questions and Exercises

1. Describe the roles of *pure experiment* and *statistical analysis in situ,* as "ideal" extremes, in scientific research and indicate times and areas where each has been predominant.
2. What kinds of order are recognized in scientific method?
3. Describe the advantages and dangers of entering on research with hypotheses of varying degrees of elaborateness. Indicate the flexibility of factor analysis in this respect.
4. Mention three regions of investigation in which factor analysis has already been effective in revealing important functional unities for application in further experimental designs.
5. What is meant by the statement, "Factor analysis might also be called factor synthesis"?
6. Define the purposes of partial correlation, multiple correlation, and the discriminant function, comparing and contrasting them with those of factor analysis.
7. Discuss the different roles of factor analysis in the very beginnings of research in a new field and in the final stages of research.
8. State and evaluate the role of factor analysis among scientific methods, having regard to four major criteria.

CHAPTER 2

Interpretation of Correlations as Clusters and Factors

Although the full implications of the previous chapter's discussion of the methodological role of factor analysis cannot be clear until factor analytic processes are understood in some detail, the importance of the general aim of seeking functionally unitary traits is obvious. The need for discovering functional unities first thrust itself upon psychologists in mental testing where the multiplication of tests of this and that alleged special ability met the opposing hypothesis that most of these tests were measuring much the same thing, namely, general intelligence. But Spearman's demonstration (**112**) in 1904 that a single factor could be found running through most mental tests, though it presented the first formal and adequate statement of factor analysis, did not constitute the first thinking in terms of operationally definable factors. As Burt shows in his search for the very earliest roots of modern factor analytic ideas (**14**), both Galton, the inventor of the correlation coefficient, and Pearson, who explored its properties, had raised the question as to the nature of the common cause responsible for any two measurements being correlated.

INTERPRETATION OF CORRELATIONS

Basically, an attempt at factor analysis is made whenever anyone tries to interpret a correlation between two things. As we know from elementary statistics, there are three possible ways of explaining an established correlation between two variables. Thus, if we take the frequently observed correlation of 0.6, among children, between reading ability and intelligence test score, we may interpret this result by saying that (a) intelligence is one of the principal causes

of good reading ability, i.e., that the individual differences in intelligence are one influence, along with other influences, producing the observed individual differences in reading ability, (b) there is some third ability or power of concentration (or maybe a group of abilities) common to the determination of both reading and intelligence test performance, and (c) good reading is itself the cause of some of the good performance in an intelligence test. Thus (c) supposes a causality which is just the reverse of (a). These three possibilities, when any two things are correlated, are represented in Diagram 2 where the shaded area represents a number of common elements, i.e., common determinants: items which can vary between positive and negative (or zero) contribution to the score on A and B.

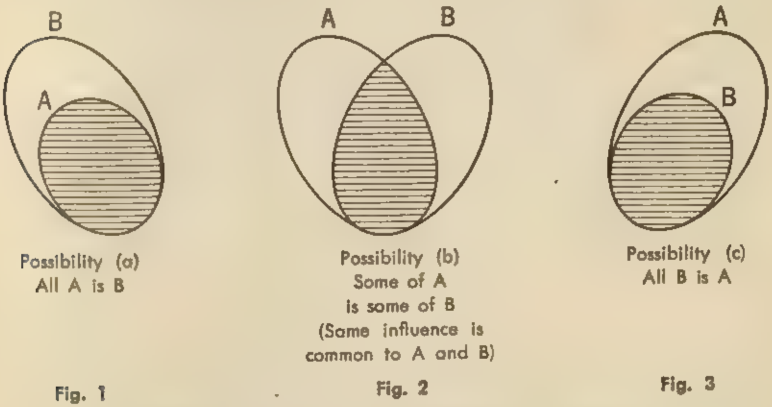


DIAGRAM 2. Interpretations of the Correlation Coefficient in Terms of Elements of Common Variance.

If we know, from ulterior information, which of these arrangements exists, we can even determine from the correlation the amount of the common elements, i.e., the amount of variance or the number of common elements that contribute to both scores. Thus, a correlation of 0.6 means in (a) that 36 percent (i.e., r^2 multiplied by a hundred[1]) of the elements in B are in A; in (b) that 60 percent (i.e., $r \times 100$) of the elements in B are in A, i.e., are common; and in (c) that *all* the elements of B are in A.

Generally, however, we have no means of peeping behind the

[1] r^2 is multiplied by a hundred simply to get rid of decimal points and to express the result as a percentage.

scenes. Nature presents us only with the correlations and we have to infer the machinery which accounts for them—this is the very meaning of scientific research. But, although the definitive interpretation of a *single* correlation coefficient is impossible, the sources of variation (factors) which account for the correlations can be made more determinate as we take more and more variables—and the many correlations among them—into our study. The student may think of this improvement in terms of the number of relations which become defined, for these increase disproportionately to the increase in number of variables. Thus there are three relations (and correlations) possible among three variables, fifteen among six, and sixty-six among twelve. As more variables and correlations are taken into account our freedom in finding alternative hypotheses as to the number and nature of the factors which could fit the observed correlations becomes increasingly restricted, for a greater variety of data gives perspective.

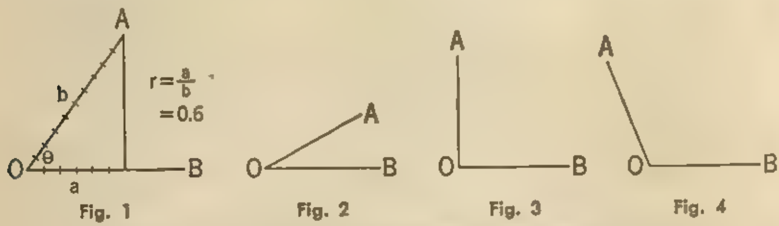


DIAGRAM 3. Correlation Coefficients Expressed as Angles.

A factor, for the present, will be defined simply as a source of variation, i.e., of individual differences, operating in two or more variables and usually spreading over performance in quite a number of variables at once. Now we can best follow the application of factor analysis to observed variation in many variables by adopting a pictorial, geometrical presentation. Experience shows that most students find this manner of symbolization far more readily intelligible than the equivalent algebraic presentation. Some of the parallel steps in the algebraic proof will be encountered later.

VECTOR REPRESENTATION

Geometrically, a correlation coefficient can be represented as an *angle between two vectors* (i.e., directed lines of definite length) constituted by the two variables in question. For the algebraic prop-

erties of cosines parallel those of correlation coefficients. Thus a correlation of $+0.6$ can be represented by this convention as shown in Fig. 1, Diagram 3, i.e., as a moderately acute angle. Why is this correlation represented as an acute angle? It is so represented by a convention that the *cosine* of the angle (see Fig. 1, Diagram 3, where cosine $\theta=\frac{a}{b}$) shall equal the observed correlation coefficient. If we follow this convention, all the ensuing calculations or drawings that we carry out in space will agree with and work out appropriate to the corresponding algebraic (implying arithmetical) transactions with correlation coefficients. No further proof of the correctness of the convention need be offered at this stage.

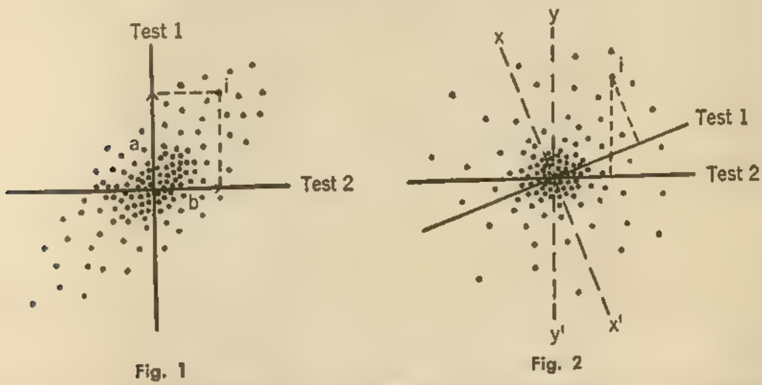


DIAGRAM 4. Two Modes of Representing a Correlation Scattergram and Individual Projections on Variables.

The student is probably accustomed in elementary statistics to showing a correlation graphically by placing points for persons on a graph formed by using the two tests as rectangular coördinates. Thus a positive correlation would be shown as in Fig. 1, Diagram 4, in which each point is a person with two scores, as illustrated by the case *i*.

By this means of representation the correlation becomes revealed through the oval form of the swarm. Instead of allowing it to reveal itself in this way, however, we could agree to bend the coördinates, test 1 and test 2, together until the swarm of points becomes circular and take the angle between these coördinates as indicative of the correlation; for the amount of rotation of test 1 and test 2

required to produce equal density of points in all directions can be precisely fixed.

If one imagines test coördinates 1 and 2 as fixed in some rubber-like substance, bending them closer together would actually *increase* the narrowness of the oval form of the points. The student's discovery that on a graph the *opposite* happens can only be seen by reference to the real meaning of projection.[2] When we talk of the projections of a point on coördinates, we mean the lengths a and b for any point such as *i* in Fig. 1. Projections in the opposite direction from the origin naturally would be negative. Now the projections of individual points in Fig. 2 are similarly obtained by dropping perpendiculars from each one onto the coördinate axes, and it will be seen that any point falling to the upper right of the dotted line xx' will have positive projections on test 1, while any point to the right of line yy' will have positive projections on test 2. Consequently, points in the area yox' will be positive on both tests, those in xoy *and* $x'oy'$ will be positive on one and negative on the other, and those in xoy' will be negative on both. Since a positive correlation means that the $++$ cases and the $--$ cases are more numerous (or larger) than the $+-$ and the $-+$ cases, it would have to be represented, when the points are distributed in an even, circular (bull's-eye) form, by a bending together of the test coördinates as has been done here, increasing the former sectors and decreasing the latter.

When this circular position (of concentric circles of points increasing in density toward the origin) is reached, the cosine of the angles between the two test coördinates equals the correlation coefficient. That is to say, the cosine of the angle 1-0-2 in Fig. 2 works out the same as the tangent of the regression lines in Fig. 1 and the same as the value obtained from putting the values for all the points in the ordinary product-moment formula for the correlation coefficient. This explains the angle convention introduced a few paragraphs above.

Actually—though the student need not know it with precision at this stage—the complete mathematical convention in mathematical terms is that correlation between any two variables is equal to the scalar product of their vectors. That is to say, both the *length* of the lines and their *directions* are taken into account, as in any true vector quantity. The correlation is thus the projection of one line upon the

[2] See also Diagram 20, Chapter 13, for projection on oblique axes.

other, as a is the projection of b upon OB in Fig. 1, Diagram 3. If b happened to be only half as long as it is, this projection (correlation) would be only half what it is, despite the angle being unchanged. However, in our first simplified account here, our measures of the variables are all taken to be in standard scores. This implies that they have equal and unit variance, so that all vectors are unit length and the correlations are directly equal to the cosines.

If the student will glance again at Diagram 3, he will recognize that in following the convention systematically a positive correlation will be shown by an acute angle and a zero correlation by two vectors at right angles, for the cosine of 90° is zero. Further, a negative correlation is represented by an obtuse angle—for by the rules of trigonometry the cosine of an angle between 90° and 180° is negative. By this means, it is possible to show at a glance the relations among quite a number of variables once their intercorrelations are known. Thus when a group of variables go together in a decided fashion, i.e., when all their intercorrelations are high and positive so that they form a correlation cluster, they will appear spatially like a quiver of arrows. This means that a person high in one of these tests is likely to be rather high in all of them, while a person who is low (negative in standard score) will have projections on the extension of the cluster of lines emerging on the opposite side of the origin.

CORRELATION CLUSTERS

In general we shall find that when a model of the correlations among many tests has been constructed, we get a spatial model of the kind shown in Diagram 5. Here the *test vectors,* as they are called, stick out through all points of the compass; but sometimes certain tests fall in clusters whereas others are isolated. In this case twelve tests are shown to fall in two clusters, of five and four (3, 4, 9, 11, and 12 in one; 1, 6, 7, and 10 in the other), with three tests remaining as isolates (2, 5, and 8). (Incidentally, the two clusters are negatively correlated since an obtuse angle exists between the lines running through each cluster.) Clusters of this kind may be found in psychology among, say, various tests of manual dexterity or in the group of trait elements that go to make up extraversion (though we have no particular evidence that the manual dexterity and the extraversion clusters are negatively correlated in the same way as are the clusters in Diagram 5). Any single cluster of this

kind we shall call a *surface trait* because all we have is an obvious indication that certain things superficially go together, though we have no knowledge as to why this is or whether it is due to a single underlying common influence or to many such influences superimposed.

A word may be necessary to clear up some questions that may rise in the reader's mind concerning the construction of Diagram 5. One can begin with test 1 and draw a line from the origin in any direction at random to represent it. The position of test 2 is then fixed by the angle it bears (as found by looking up the angle with a cosine equal to its correlation with 1) to test 1 and so on with the other tests. The coördinate system provided by axes OF_1 and OF_2 is put on the model afterwards and for convenience is arranged, as usual, with the axes parallel to the edges of this page. *The relation of the test vectors (which constitute a rigid system among themselves) to the coördinates is thus quite loose and arbitrary, since we could have begun by drawing the vector for test 1 anywhere.*

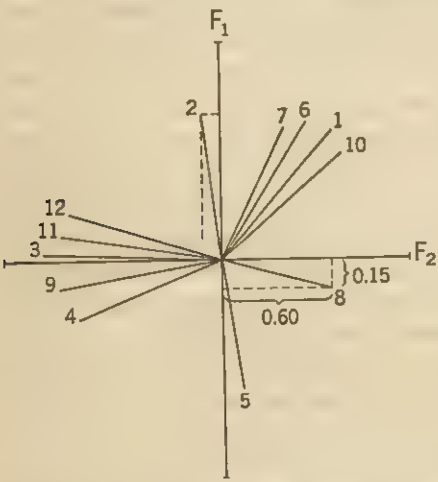


DIAGRAM 5. Clusters and Factors Among Variables as Vectors.

FACTORS AS COÖRDINATE AXES

Let us overlook the fact that the coördinate system can apparently be spun around like the bars on a roulette wheel to any position, and let us assume for the moment (as will later be demonstrated) that there is no difficulty in agreeing that one rotation position is more meaningful than any other. It then becomes possible to point out very simply what the factor analysis of a set of correlations among variables is really doing. What we call factors are essentially nothing more than axes and the factor analyst has merely imposed *a framework of coördinate axes* upon the structure of test vectors built up

from the correlations, as shown by the two dotted lines F_1 (factor 1) and F_2 (factor 2) in Diagram 5 which represents a case where only two factors are involved. By this means any one test can be represented by the numerical value of its projections on the two coördinates, as test 8, for example, is resolvable into -0.15 of F_1 and $+0.60$ of F_2.

In this way, all twelve tests can be represented in terms of only two factors. This has theoretical advantages in that we can begin a search for two hypothetical powers or tendencies which lie behind the performances in all twelve tests, and it has practical advantages in that we may hope to substitute two tests to tell us practically all that is now being achieved by the use of a long battery of twelve tests.

As an example of this latter advantage—economy of measurement—we may note that test 3 in Diagram 5 measures practically all that tests 4, 9, 11, and 12 measure and is a practically pure measure of factor 2 (but in the negative direction). Similarly we could measure pure factor 1 by finding some test midway between 2 and 7, and in default of such a discovery, we can obtain a tolerably good measure of it by combining (averaging) the scores on 2 and 7. Thus each individual person would be assigned scores on two factors instead of scores on twelve tests, and these two scores would tell us almost as much about his behavior as the original twelve. In fact, if we know the projections of all twelve tests on the two factors, we can make fairly good estimates from his known *two-factor scores* as to what his performances will be in *each* of the tests.

SPECIFICATION EQUATION

This estimation process will be taken up in detail when the factor analysis itself has become more clearly understood; but even at this stage an example may help clarify the essential relation of factors to variables. The relation of test 1 to the factors is given by projections (or loadings as they are called in factor analysis) of -0.15 and $+0.60$ as shown above. If performance in the test and endowment in the factors is given in standard scores, this means (as will be seen later) that we can predict an individual's performance by the following specification equation:

$$\text{Performance in test } 1 = -0.15F_1 + 0.60F_2$$

where F_1 and F_2 are the individual's endowments in F_1 and F_2 (which we should have to know from other sources). Or if we want to

speak in general terms referring to no particular individual, we should say that the *variance* in test 1 is built up by a contribution to the extent of $(0.15)^2$ from the variance of the first factor and to the extent of $(0.60)^2$ from the variance associated with the second factor. The reader will probably not need to be reminded that the variance is the square of the standard deviation. The statement that 36% of the variance in this performance (test 1) is associated with Factor 2 means, among other things, that if all individual differences—all variance—in Factor 2 were abolished, the variance in test 1 would be reduced by 36%. Thus if 25% of the variance in stature in sons is associated with variance in the stature of fathers (corresponding to a father-son stature correlation of +0.5), we should find, on taking the variance of measured stature among sons all having fathers of the same height, that it is reduced 25% relative to that of the general population.

The sign of the factor loading in the above specification equation indicates the direction in which the factor operates. A greater endowment in Factor 2 gives an increase in performance but a greater endowment in Factor 1 tends to reduce one's performance in test 1. The variance, which is the square of the correlation (loading), of course has no sign, and the total variance breaks down into the sum of the contributory variances regardless of the signs of the correlations.

Just as we have read off from the projections on Diagram 5 the loadings for test 1, so we can read for further test vectors:

$$\text{Performance on test } 2 = 0.80F_1 - 0.10F_2$$
$$\text{Performance on test } 3 = 0.01F_1 - 0.99F_2$$

and so on. Thus each test performance resolves into a function of the same two factors. The particular test performance of a given individual can then be estimated from his particular endowments in those factors.

At this point, however, seeing the comparatively well-defined grouping of variables in Diagram 5, someone may ask if it might not be more convenient to deal with clusters instead of factors. Many psychologists—to judge by the greater frequency with which cluster rather than factor analysis has been used in the past—felt that they had their feet more firmly on the ground when they dealt with actual clusters instead of the more shadowy factors which may be abstracted

from them. Besides, it *seems* so much easier to inspect the correlations for clusters than to go through the rather prolonged calculations which, as we shall see, become necessary in extracting factors.

But these are illusory attractions. In the first place, there are generally many more clusters than factors. In the personality sphere, for example, we can represent individual differences in some two hundred personality variables either by sixty clusters or by about twelve factors. Factors are more economical. Second, clusters may be highly correlated among themselves so that they do not offer independent coördinate axes by which the tests that do not fall in clusters (the numerous isolates) can be brought into a common scheme of representation. Third—and this is the real Achilles' heel of the cluster—the level of mean intercorrelation by which we limit admission of tests to a cluster is arbitrary. Some people consider that only those tests which correlate together above +0.8 belong to the same cluster, whereas others would put it as low as +0.3. In spatial representation this means that the fan of vectors is taken in some cases to spread very widely and in others narrowly, in an arbitrary fashion which eventually causes confusion. Indeed, a last and most disabling difficulty with cluster analysis is that in any real data the clusters tend to straggle and run one into another, like clouds in a stormy sky, so that any separation of functional unities by this means becomes very arbitrary and undependable.

Questions and Exercises

1. Explain three possible interpretations that could be given to a correlation of +0.50 between two variables, X and Y.
2. Draw a rough sketch showing points for twenty people on a correlation plot and representing what you would estimate to be a correlation of about +0.5 between the two variables concerned (a) in the form of an ordinary scattergram with variables as rectangular coördinates and (b) in the form of uniform-density concentric rings with coördinates at appropriate angles.
3. Draw the angles between two test vectors of unit length corresponding to tests having the following correlations: +0.00, +0.20, +0.50, +0.71, +0.87, +1.00, −0.50, −0.71, −0.87, −1.00.
4. By observing the positions of the two test vectors in each drawing of problem 2, formulate a general rule regarding the size of the angle between them as their correlation increases from −1.00 to 1.00.
5. Plot the test vectors represented by the following table, assuming that the factors F_1 and F_2, which have been chosen as the coördinate axes, are mutually perpendicular. If F_1 and F_2 were known to be correlated

positively with each other, how would this change the graph? In what way would this change affect the projections of the test vectors upon F_1 and F_2?

	F_1	F_2		F_1	F_2
T_1	.85	.53	T_{11}	−.45	−.89
T_2	.57	.82	T_{12}	.68	.73
T_3	−.19	−.98	T_{13}	−.80	.60
T_4	−.97	.25	T_{14}	−.76	.65
T_5	.63	.78	T_{15}	−.64	−.77
T_6	−.83	−.56	T_{16}	−.98	−.19
T_7	.94	−.35	T_{17}	.87	.50
T_8	−.31	−.95	T_{18}	.43	−.90
T_9	.91	.42	T_{19}	−.54	−.84
T_{10}	−.99	.14	T_{20}	−.96	−.28

6. By examining the graph in problem 5, determine whether the clusters of test vectors are positively or negatively correlated. Is it possible to determine the exact number of clusters in this group? How does this graph illustrate the possible fallacy of using clusters rather than single factors to describe the tests?
7. Which test or tests might best be used to represent each cluster in problem 5?
8. State four major difficulties in the attempt to simplify the description of many variables by means of cluster (or surface traits) rather than by factors.

CHAPTER 3

On Obtaining Factors from a Correlation Matrix

The previous chapter has indicated essentially what factors *mean*; but it has not shown how we arrive at them in practice, for it is easy to see that the graphical method described is impracticably rough, and indeed impossible for more than three factors (three dimensions). There must be some method of finding factor loadings by calculation carried to any required degree of accuracy. Moreover, our first discussion has not shown how we decide on a *particular position* for the coördinate axes. Furthermore, it has proceeded as if all sets of intercorrelations could be resolved into two factors—the two coördinate axes drawn on paper—whereas in a sufficiently large set of variables it must often happen that decidedly more than two general factors or influences are at work to account for the peculiarities of the observed covariation.

NUMBER OF FACTORS

This problem of deciding *how many* factors are at work needs to be solved before the others, and we shall first approach it in terms of the geometrical mode of representation already adopted. Actually, the reason for being able to operate with two coördinate axes only, in the previous chapter, is that we took a special example—one in which the angles among test vectors happened to be such that the vectors could be fitted into two dimensions. They would thus lie flat in the plane of the paper. Only specially chosen correlation coefficients would yield this result.

But in any real case of correlations among several variables, taken at random from research results, it is quite unlikely that the correla-

tions will give angles that will do just this. Thus the angles drawn in Diagram 6, which represents a more typical case than our previous example, have cosines equal to the correlations found between two intelligence tests (Y *and* Z) *and a test of mechanical aptitude* (X). If these were cut out in paper, as sectors of a circle with a fixed radius, the student would find (he may care to try) that he encounters something more puzzling than was encountered among the correlations of Diagram 5. The vectors refuse to lie in a flat plane. When we fit the pieces together, the fan of vectors now stands up in three dimensions like the spokes of a partially opened umbrella. Indeed it will be seen that to get O, X, Y, and Z in a single plane, as in Diagram 5, it would be necessary for the angle XOY plus the angle YOZ exactly to equal angle XOZ (or the alternative combination) and this special condition will obviously not often be met in nature.

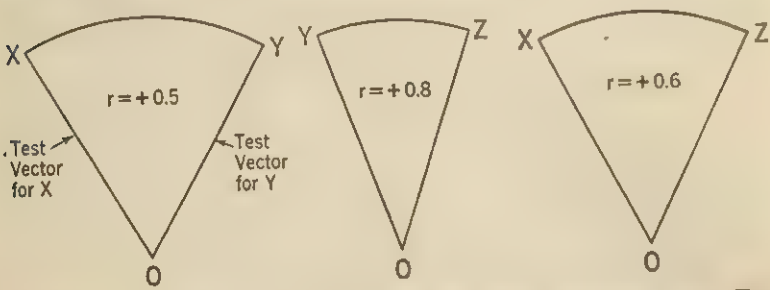


DIAGRAM 6. Three Variable Correlations not Resolvable into Two Dimensions.

In the situation represented in Diagram 6 the points X, Y, and Z (and the test vectors spreading from O which they demarcate) now lie in ordinary three-dimensional space and require *three* coördinate axes on which to fix their positions. That is to say, each of the test vectors now has projections or loadings on three factors, and the specification equation for any one would run as follows:

$$\text{Performance on } X = s_1F_1 + s_2F_2 + s_3F_3$$

Almost any sufficiently varied collection of psychological or sociological data will, however, yield correlations among the variables which cannot even be fitted into a *three*-dimensional model. What happens in such a case, where four or more influences are at work,

to the form of our model? It can be set up only by using four or more dimensions, and this can be handled in a physical sense only by holding some dimensions constant while we construct a model in the remainder.

HYPERSPACE

But we need not depend any longer upon models. Mathematicians have developed methods of handling geometrical problems in four, five, and higher dimensional space even though we can no longer visualize or construct models in such space. Since correlations requiring more than three factors to explain them are quite frequent, such methods are important. This space beyond three dimensions is called *hyperspace* and it is, of course, purely imaginary and used only to symbolize relationships, such as the present ones, in which spatial representation is in fact symbolic from the very beginning. Fortunately, the geometrical problems of hyperspace can be worked out by sticking to the same rules as apply to our familiar three-dimensional space, and there is no need to hold one's breath or lose orientation when the correlations provided by experiment show that we have to go beyond the frontiers of visualizable room. In fact, as indicated above, problems of hyperspace can be handled visually by taking two or three dimensions at a time, observing the projections on these and neglecting the rest for the time being. The student interested in pursuing further the application of geometrical principles to correlation, in its basic theoretical aspects, should read the article by Jackson (**79**) cited below.

What factor analysis means by factors is therefore nothing more than the dimensions (independent coördinate axes) of the space required to contain a certain set of correlations when they are spatially represented. If the correlations given by nature are of such relative magnitudes that the vectors representing the psychological variables can all be represented in the flat, as in Diagram 5, it means that the covariation manifested in these variables can all be expressed by only two truly independent sources of variation. If, on the other hand, the given correlations force us into four dimensions, we know that there are four varieties of variation—perhaps variations in four distinct kinds of ability—required to account for the individual differences of performance in the dozen or so tests used.

FACTOR LOADINGS

It is now time to turn to the algebraic statement of the problem in order that we may show how the factor loadings, i.e., the projection of the test vectors on the coördinate axes, are actually found. For in practice we do not proceed by building up a physical model of the test vector structure and then putting coördinates through it. Indeed as we have seen we might be forced by this approach into the Alice-in-Wonderland game of trying to build a model in four dimensions! Instead we make *calculations which give us straight away the projections of the vectors on the factor coördinates* (axes). Then when we want a visual picture of the relationships, we can take any two dimensions at a time and draw upon an ordinary graph the positions of the vectors relative to these particular axes, using the projections given by the initial calculations.

Before describing the algebraic process by which factor loadings are calculated from the given correlations, we must clarify a certain basic proposition about the relation of the factor loadings of two variables to their mutual correlation. If a variable *a* correlates with some component (constituent) within *a* (which we will call *F* or common factor) by the amount of, say, +0.4, and if another variable *b* correlates with the same *F* (therefore also appearing in *b*) to the extent of +0.3, then the correlation of *a* with *b* will be the product of these separate correlations, namely, +0.12. This can be shown by a geometrical proof using the above convention, but it must be noted that it does not hold for the correlation of any two variables *a* and *b* with a third, *c*. It holds only for orthogonal factors and where the two variables have nothing in common but the one factor.

It will be seen that in this argument we are dealing with a situation which may be illustrated by the typical case of test measures *a* and *b* in Fig. 2, Diagram 2. That is to say, they have the same factor in common, but they differ from that figure in so far as one measure, *a*, has more of the factor than *b*. This is the most general possible assumption about the relations, and Figs. 1 and 3 may be regarded simply as extreme instances thereof. Expressed in the general form this statement runs:

$$r_{ab}=r_{aF}\times r_{bF} \tag{1}$$

when *a* and *b* have nothing in common but *F*.

Thus if success in arithmetical performance depends on intelligence (g) to the extent of a correlation, r_{ag}, of $+0.6$ between arithmetic and intelligence tests, and if success in boxing depends on intelligence to the extent of a correlation of $+0.2$, and if arithmetic and boxing have nothing else in common, we should find that the correlation between proficiency in arithmetic and boxing would be $(+0.6)\times(+0.2)=+0.12$.

By extension, if two tests have *more* than one factor in common, say F_1, F_2, and F_3, it can be shown that their correlation r_{ab} can be calculated by:

$$r_{ab}=(r_{aF_1})(r_{bF_1})+(r_{aF_2})(r_{bF_2})+(r_{aF_3})(r_{bF_3}) \tag{2}$$

That is to say, the correlation of two variables is the sum of the products of their corresponding loadings in the common factors, providing the factors are independent of one another. However, the task we actually face in psychological research is not this process of obtaining the correlations from known factor loadings, which is what the present insight into structure so far permits us to do, but the converse process of *obtaining the factor loadings from known correlations.* Let us return to consider this new problem in the simplified situation of only one factor and let us suppose that we have the correlations of test a with ten other tests b, c, d, e, f, g, h, i, j, and k each possessing some amount of the same factor. Then:

$$\begin{aligned} r_{ab}&=(r_{aF})(r_{bF})\\ r_{ac}&=(r_{aF})(r_{cF})\\ r_{ad}&=(r_{aF})(r_{dF})\\ &\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\\ &\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\\ r_{ak}&=(r_{aF})(r_{kF}) \end{aligned}$$

It will be noticed that though a new correlation coefficient enters at the right in each new row, one correlation—r_{aF}—is repeated. Consequently, if we summed a very long column of such products, we could expect the differences in the specific correlations (r's of b, c, d, etc. with F) at the right to some extent to cancel out in any comparison of the column totals, leaving us with a figure which represents, in this case, the sum of ten r_{aF} coefficients. In other words,[1] *the mean correlation of test a with all other tests (obtained from this sum) is*

[1] To make this mean from the column total *equal* to r_{aF}, we need to divide it by the mean r of all other tests in the column with the factor.

closely proportional to its correlation with the factor, r_{aF}. In fact, as we shall see later, we obtain this correlation by summing the column and dividing the total by the square root of the sum of all columns.

From this we can see why the essential process in factor analysis, namely the extraction of factors from correlations, takes the form now to be described. This formal process will first be illustrated by a simple arithmetical example in which we take intercorrelations among eight tests and try to find how much each test has of the common factor possessed by all of them.

At this point we may pause to notice that as a result of the above requirement in computing r_{aF}, a factor analysis cannot begin until *every possible correlation* among the variables in question has been worked out. This means that every person in the population measured must be measured on every test. The tests are then correlated in pairs, systematically for all possible pairs, using whatever form of the correlation coefficient is appropriate. The systematic procedure would correlate test 1 with tests 2, 3, 4, and so on to test n, producing the coefficients in the first column of Table 1 below. Then variable 2 is correlated with tests 3, 4, etc. on to test n, and so on for the remaining columns.

THE CORRELATION MATRIX

A simple application of the formula for the number of combinations of two things from among n things will show that there are always $\frac{n(n-1)}{2}$ coefficients to be worked out from n variables in order to get all relations fixed. This can be seen readily also by drawing up what is called a *correlation matrix*, as in Table 1, in which the n variables are arranged along the top and along the side of a square, and the correlation obtained between any two variables is placed in the cell formed by the intersection of row and column. Applying the formula to the case of 8 variables as shown below, we obtain $\frac{8 \times 7}{2}$, i.e., twenty-eight coefficients. However, these twenty-eight, which would occupy the lower left of the matrix, have been repeated in the upper right of the matrix. For although the triangular lower left is sufficient to represent *all* correlations, it will later be found convenient for many purposes to have the correlations in duplicate. For example, when we need to add up all the r's (as we shall henceforth describe

TABLE 1. Correlation Matrix of Eight Variables

Test Variables	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	Repeating Columns
V_1	(30)	20	28	36	−09	01	06	06	1
V_2	20	(30)	00	52	36	−04	04	04	2
V_3	28	00	(30)	04	−36	04	04	04	3
V_4	36	52	04	(90)	66	−02	43	39	4
V_5	−09	36	−36	66	(90)	−06	27	24	5
V_6	01	−04	04	−02	−06	(05)	09	08	6
V_7	06	04	04	43	27	09	(80)	72	7
V_8	06	04	04	39	24	08	72	(60)	8
Σr	1.18	1.42	.38	3.28	1.92	.15	2.45	2.17	T = sum of $\Sigma r = 12.95$
$\frac{1}{\sqrt{T}}(\Sigma r)$	.33	.39	.11	.91	.53	.04	.68	.60	$\sqrt{T} = \sqrt{12.95} = 3.60$
	.3269	.3933	.1052	.9086	.5376	.0416	.6787	.6011	$m = \frac{1}{\sqrt{T}} = \frac{1}{3.60} = 0.277$[a]

[a] The method by which the communalities are "guessed" for the diagonal will be described later.

correlations coefficients, for brevity) of, say, test 4, it is easier to do so in a column than to follow the elbow of a broken column and row.

To find the amount of the general factor in test 1, we should, according to the principles just announced, add up all the correlations in the first column and take the mean. There are seven such correlations but, for reasons given later, an eighth is added—one in each column—to fill in the blank spaces represented by the diagonal from upper left to lower right.

This additional r in each column[2] is called the *communality* and represents the correlation of the test with itself *in so far as this is due to the common factor or factors.* The communality is thus equivalent to the other r's in the column which, incidentally, also represent the relation between two tests so far as it is due to common factors. But the communality is *not* to be confused with the *reliability coefficient* of the test in question. For the latter represents the r of a test with itself due to the factors unique to the test *as well* as to those shared with other tests. The communalities are not

[2] The reader is reminded that r will henceforth be the shorthand for correlation coefficient.

given by the experiment—one begins with blanks along the diagonal—and they have to be estimated or guessed at the beginning of the computation. The particular principles by which this estimation of communalities is carried out will be explained later, but to remind us from the beginning that they are always only estimated and inexact, we have put in *rounded* numbers along the diagonal. Having completed all the cells in the matrix by the estimation of communalities, we then proceed in the factorization by adding up each of the columns (attending to signs as in algebraic addition generally), and we put the totals in a row at the bottom labelled Σr (Σ meaning the sum of —pronounced sigma). Instead of dividing these totals by eight, however, in order to get the simple arithmetic mean, we divide them all by a value $\sqrt{T}$, where T is the total of the eight column totals. The reason for this mode of reducing the totals to comparable values lies in the need defined in footnote[1] for correcting each column total for the effect of multiplying r_{aF} by each of the other r's in turn (see page 39).

THE CENTROID

If we turn back to the corresponding geometrical presentation to see how the above algebraic process has located the length and direction of the first factor axis, we see that it has in fact put the axis through the center of gravity of the swarm of points constituted by the ends of the test vectors. If we imagine the ends of all the test vectors to be small lead balls and the vectors themselves to be weightless threads, then the centroid position, marked X in Diagram 7, would be the point about which all these weights would balance. A vector drawn from the origin through this point therefore represents *the* direction of variance common to all these test vectors.

The length of this axis from O to X, moreover, represents the amount of the variance from the origin which lies on this direction, i.e., in this first factor. The value $\sqrt{T}$ which we used above is in fact the length of this axis, and by dividing the test projections by it, we express them relative to an axis of *unit* length. If for a moment we let the geometrical exploration run a step ahead of the algebraic, we may notice that the extent of the need for extracting a *second* factor is indicated by the extent of the scatter of these points about the center of gravity just located. Shifting the origin

from O to X has left the test vectors much shortened, but there is still some variance in them, about X, to be accounted for by further factors, the algebraic extraction of which will soon be described.

Incidentally, the term centroid method is generally attached to those developments of factor analysis associated with the name of Thurstone (**126**) since he was largely responsible for introducing it.

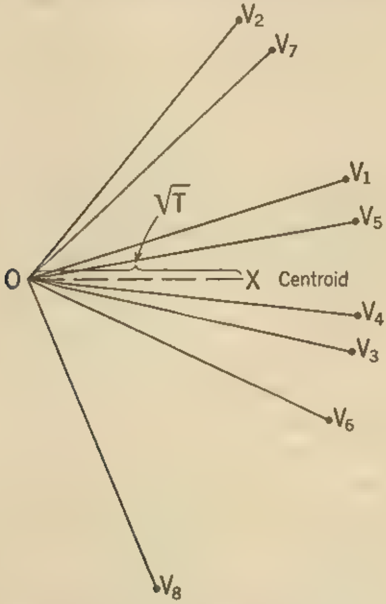


DIAGRAM 7. Calculation of the Centroid for a Set of Correlating Variables.

But the student should recognize that its soundness does not stand or fall with all the rest of Thurstone's techniques of factorization. It is a perfectly general method of finding the factors which can later be handled by the special rotation methods (see below) of Thurstone, or by other methods.

The results of division by $\sqrt{T}$ are given correct to two decimal places[3] in the second row at the foot of the matrix (Table 1). These are the required factor loadings or saturations for the first factor.

[3] The initial values from which these are rounded are shown in the bottom column.

They represent the projections of the points at the end of the test vectors upon the *first* coördinate axis. The way in which these obtained first factor loadings are used in the steps to obtain the loadings for the *second* factor is explained in the next chapter.

Meanwhile the practical-minded reader concerned about research organization may be interested to reflect that the factor analytic approach, by requiring every possible correlation to be worked out among variables in order to form the matrix and by requiring, as will be seen later, that a sufficient population of variables be included in the matrix, makes rather heavy demands both on data gathering and on the organization of statistical computing. However, when these difficulties are overcome, the method yields answers about the nature and roles of influences which are unobtainable by any accumulation of less organized studies, e.g., those taking variables two or three at a time and working out only some prejudged important relations among them.

Since the time, thought, and money that has gone into these numerous local studies far exceeds that which has so far gone into properly designed and organized factorizations, it is unfortunate that the material scattered in the former cannot be put together in a single complete matrix. It would be a fine task of salvage to rescue individual correlations from many years of publications, in a skilled jigsaw puzzle enterprise, to build up large correlation matrices from which overall perspectives about factors could now be obtained. But it is not possible with any simplicity or accuracy to combine in one matrix correlations obtained upon different population samples. Consequently the many studies investigating the relations of only two or three variables need to be done again in well-chosen larger companies of variables if that illumination of mechanisms which is possible through the simultaneous consideration of many relations, and which is made possible by factor analysis, is to be achieved.

Questions and Exercises

1. Under what conditions will the correlations among three or more variables yield angles such that the variable vectors can be represented in a single plane?
2. What is meant by hyperspace and what geometrical rules apply to it?
3. What is the meaning of factor and factor loading in the geometrical representation of correlations?
4. How many correlation coefficients would need to be worked out in the correlation matrix for nine variables?

5. Why is the process of factor extraction which was popularized by Thurstone, called the centroid method?
6. Find the correlation of test A with test B due to the following correlations of each with the four different factors shown in the table. Note that both tests may be highly correlated with a given factor while their correlation together may be quite low, and vice versa.

Test	Loadings on			
	F_1	F_2	F_3	F_4
A_1	.3	−.1	.9	.9
B_1	.7	.4	.0	−.2
A_2	.3	−.7	.9	−.6
B_2	−.2	.4	.8	.7
A_3	.5	.2	.4	.6
B_3	.3	.4	.5	.5
A_4	.0	−.7	.8	.5
B_4	−.2	−.6	.5	.2
A_5	−.6	−.2	.9	−.4
B_5	.3	−.4	−.7	.1

7.

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9
t_1	(.70)	.62	.75	.57	−.19	.54	.25	−.77	.11
t_2	.62	(.80)	.84	.48	−.32	.50	−.19	.34	−.53
t_3	.75	.84	(.80)	−.48	.34	−.45	−.22	.38	.40
t_4	.57	.48	−.48	(.70)	−.06	−.06	−.47	.43	.10
t_5	−.19	−.32	.34	−.06	(.50)	.06	.47	.49	.76
t_6	.54	.50	−.45	−.06	.06	(.60)	−.16	.19	.10
t_7	.25	−.19	−.22	−.47	.47	−.16	(.20)	.04	.38
t_8	−.77	.34	.38	.43	.49	.19	.04	(.50)	.67
t_9	.11	−.53	.40	.10	.76	.10	.38	.67	(.40)

From the correlation table above determine the mean correlation of each test with the nine tests of the group and calculate the corresponding factor loadings for each test using the method of Table 1, page 41. Observe that all figures in the computation should be rounded off to two decimal places since the correlations are accurate to only two places. Also, it is often easier to multiply the sum of each column by the number $1/\sqrt{T}$ rather than to divide each by $\sqrt{T}$.

8. What is the difference in meaning between a communality and a reliability coefficient?

CHAPTER 4

Extraction of Successive Factors

The most frequently used and basic method for extracting a single factor has been briefly indicated. Questions now arise as to how we repeat this process for further factors and how we know when *all* the factors in a set of correlations have been extracted. In the geometrical representation, as seen above, the number of factors is the number of dimensions required to contain the correlation structure, but what is this in terms of the algebraic process and the arithmetical computations which follow from it?

In answering this, we follow in the tracks of the historical approach to factor analysis. Let us consider the actual line of reasoning by which Charles Spearman (**112, 113**) in 1904 developed the first theorems in factor analysis when he was attempting to understand the nature of intelligence as a single general factor among all tests of cognitive ability.

COMMON FACTORS

As courses in general or elementary statistics point out, the correlation that exists between two abilities when each is measured by a good test, i.e., a test of high reliability, is generally higher than when each is measured by a poor test (or in poor conditions) yielding unsatisfactory reliability coefficients. The r between them, we say, is *attenuated* in the latter circumstances by the chance errors in each test. For a significant correlation represents some kind of *order*, and any kind of disorder from errors of one kind or another will therefore tend to act *only in the direction of reducing the correlation.* In this respect a correlation coefficient is to be contrasted with a mean and most other statistics; for there the effect of chance error is to make the figure more uncertain, i.e., to increase dispersion but not systematically to alter the mean in one direction.

A correction for this attenuation of r has long been known, which succeeds in indicating what the r *could* be if errors were eliminated. It runs as follows:

$$r_{ab} = \frac{\sqrt[4]{(r_{a'b'})(r_{a'b''})(r_{a''b'})(r_{a''b''})}}{\sqrt{(r_{a'a''})(r_{b'b''})}} \tag{3}$$

where a' and a'' are two actual tests (test and retest) measuring, with some error, the ability a; while b' and b'' act similarly with respect to b. It will be noticed that we have all the possible combinations of the actual measures correlated in the numerator and the two reliability coefficients in the denominator. The better the reliability coefficients the nearer the observed r's will be to the true r.

Now Spearman's novel way of looking at this formula was to see it as a statement for obtaining the correlation that would exist between two tests of the same common factor—which he called g or general ability—*were it not for the specific factor which each also has in itself alone*. That is, he perceived that whatever is specific to one given test behaves just like a chance error (additional to the chance error of measurement) as far as the measurement of the common factor in it is concerned. Indeed we can regard any four tests, 1, 2, 3, and 4, which we suspect of having a general factor in common as if they were a', a'', b', and b'' in the equation above, i.e., different attempts to measure the same thing. Thus the equation enables us to find out about this same thing, i.e., to discover how much general factor each has.

Spearman's new perception of the situation is best expressed by rewriting the above equation in its new characters. The four testings a', a'', b', and b'' now become four distinct tests, 1, 2, 3, and 4, which are chosen on the assumption that they are different attempts to measure the same thing, i.e., they have one factor in common which can be called g. Now the *true* correlation of a with b in the attenuation equation (3) above becomes r_{gg}, since by our assumption the g in 1 and 2 is the same factor as the g in 3 and 4. The equation then becomes

$$r_{gg} = \frac{\sqrt[4]{(r_{13})(r_{23})(r_{14})(r_{24})}}{\sqrt{(r_{12})(r_{34})}} \tag{4}$$

if we choose (arbitrarily) to take r_{12} and r_{34} as the equivalent of the reliability coefficients. (Alternatively we could take r_{13} and r_{24} or r_{14}

and r_{23}.) But r_{gg}, the correlation of the factor with itself, *is by hypothesis perfect, i.e., unity*. If we put unity in the equation, we can obtain by algebraic rearrangement[1] that

$$\frac{r_{13}}{r_{23}}=\frac{r_{14}}{r_{24}} \tag{5}$$

and that this can be continued for any number of tests in the matrix, thus:

$$=\frac{r_{15}}{r_{25}}=\frac{r_{16}}{r_{26}}=\frac{r_{17}}{r_{27}} \text{ etc.} \tag{6}$$

Spearman put (5) in the form

$$(r_{13})(r_{24})-(r_{14})(r_{23})=0 \tag{7}$$

which is the renowned *tetrad difference equation* that did signal service for years as a device for testing the number of factors, until multifactor analysis generalized the idea. It will be seen from the starting point in the formula for correction for attenuation that if the tetrad equation holds, i.e., if the given r's inserted in the equation *do* actually equal zero, then the tests are measuring nothing but *one* general factor common to all, plus certain specific factors or errors *peculiar to each test* (since a correction for attenuation in these circumstances should raise the correlation between tests to unity).

HIERARCHICAL ARRANGEMENT

Consequently, Spearman's method of examining a correlation matrix to see whether a single general factor could be supposed to operate in *all* the variables consisted in taking those variables four at a time (in every possible combination) and examining the difference, as in equation (7), to see that all were zero. From the form of this equation in (6) it will be seen that it also implies that the r's for test 1 should be proportional to the corresponding r's for test 2 and so on. In short, it should happen that when we sort the tests in order of magnitude of their correlation with test 1, they will also turn out to be in correct order of magnitude of their r's with test 2. This alternative statement of the tetrad difference is known as the principle

[1] This arrangement cannot be made by a direct transformation from (4) but involves cancellations etc. among the three alternative ways of writing (4).

TABLE 2. Correlation Matrices with Variables in Hierarchical Order

	V_6	V_1	V_4	V_5	V_3	V_2	V_8	V_7
V_6								
V_1	.90							
V_4	.82	.75						
V_5	.73	.61	.58					
V_3	.51	.49	.43	.36				
V_2	.43	.30	.25	.22	.18			
V_8	.31	.27	.21	.15	.11	.09		
V_7	.24	.15	.12	.10	.08	.06	.05	

of *hierarchical arrangement*, for it permits the arrangement of the whole matrix in a hierarchy as shown in Table 2 in which all adjoining columns are proportional.[2]

Spearman used the tetrad difference criterion only to eliminate from his battery any tests which brought more than a single general factor and broke the hierarchy. (When two variables have in common some extra group factor which they share in addition to that general factor which they share with all others, the correlation between them becomes too high for them to fit into a tetrad equation and yield zero difference.) But in the late twenties of this century, Thurstone perceived in it a device for testing how many *additional* factors might be present in the battery if we do *not* wish to purify it of further factors by eliminating variables as Spearman would have done. For concentration on a single factor he substituted the goal of seeking as many factors as might exist, and thus inaugurated multifactor analysis.

[2] It will be noticed that the *variables* (at top and side) are not in numerical order in the matrix since they have had to be rearranged from the initial order given them in the experiment in order that the *r*'s may fall in hierarchical arrangements.

Tryon (128) has generalized this system of looking for tests with similarity of correlation profile (for it would be a similarity of pattern if we did not arrange them in descending hierarchical order) into a system which he called cluster analysis. It is a hybrid system in that it requires the variables put together to have some degree of similarity of profile as well as some mutual correlation as in our earlier definition of a cluster (page 29). The element of subjective judgment in combining these requirements as well as the failure of the system to yield exact predictive equations do not justify its being studied further here, though it requires less calculation than factor analysis and is useful as a system of classifying variables.

THE RANK OF A MATRIX

The student who does not know matrix algebra would do best at this point to take it on trust that mathematicians test what they call the rank of a matrix by multiplying out all its tetrads to see whether they each have value zero. If they do, the matrix is of rank one, i.e., it contains one general factor, as Spearman's tetrad proof showed. If at least one of the tetrads does *not* come to zero (within the limits of error) the matrix is at least of rank two. The mathematician would then take tetrads among tetrads, i.e., treat the value of each tetrad as an entry in a set of four values similarly obtained; and if they came to zero, the matrix would be said to be no more than of rank three.

In general to test for higher rank in a matrix we can, for some ranks, continue to take tetrads among tetrads of tetrads, etc. We can, however, extend this idea of tetrads to include sums and differences of more than two products, by use of determinants within the matrix; this idea will be developed further in Chapter 13. Thurstone perceived that the *rank of a matrix*—when the correlations are the entries in the matrix—*is the same thing as the number of independent dimensions when the correlation structure is graphically represented* in space.

MULTIFACTOR ANALYSIS

Thus multifactor analysis—the extraction of several common factors from a set of correlations—was born. It abolished the limitation of having to explore abilities or temperament traits one at a time, by repeated experiments designed to exclude tests which broke the hierarchy of each single common factor. It enabled the psychologist, previously earthbound to a pedestrian investigation of one plane at a time, to explore many dimensions of psychological variation at once and thus to see the relation of factors one to another at a single glance.[3]

The above description of the tetrad difference and its meaning has been introduced to give some idea of the main points in the historical development of factor analysis. Today tetrad differences have meaning only as a device for testing the rank of a correlation matrix, i.e., for showing how many factors will need to be extracted before the

[3] It is an instructive comment on our habits of reading and understanding of past work that a clear statement of multifactor analysis in spatial terms was first given by Maxwell Garnett (57) in 1919 and was generally recognized and recalled only after Thurstone had made his independent discovery in 1931.

actual work of extraction begins. They do not enter into regular factor analytic procedures. Even in respect to determining the number of factors to extract, this criterion has largely been dropped in favor of other tests of the completeness of factor extraction, to be described later. For this method of predetermining the rank of a matrix is very laborious, and in practice the great majority of factor analysts simply proceed to extract factor after factor until they have evidence that nothing remains to be extracted. These tests by exhaustion can be understood after we have described the process of further extraction.

THE FACTOR MATRIX

With the above preliminary digression completed in regard to the historical transition from single factor to multifactor extraction problems and with the understanding thus gained of the characteristics of a correlation matrix containing only one factor, we return to the particular factor extraction problem started in the previous chapter. A glance at page 41 will remind the reader that we have extracted one factor from this 8 by 8 matrix and have obtained a set of *first factor loadings* which may be set out separately from the correlation matrix in the beginnings of a *factor matrix* as follows:

TABLE 3. Table of Factor Loadings (Beginning of "Factor Matrix")

Test variable	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8
Factor loading	.33	.39	.11	.91	.53	.04	.68	.60

Now, as formula (2) shows us, the correlation of two variables with each other as the result of their sharing a common factor is the product of their correlations or loadings with the latter. Thus the r between V_1 and V_2 due to the *first* factor is the product of their correlations or loadings therein, i.e.,

$$r_{12} = r_{1F_1} \times r_{2F_1} = 0.33 \times 0.39 = 0.13 \tag{8}$$

But the *observed* correlation of V_1 with V_2 is actually 0.20 (see correlation matrix, Table 1, page 41). Consequently, a correlation of 0.07 remains to be accounted for, and this must be due to *further factors that V_1 and V_2 have in common*. To see whether second and further factors are required to account for the correlations among *all* the variables therefore, we need to carry out this subtraction from the original r for every pair of variables. This is done systematically

by setting up what is called a *factor product matrix*, as shown in Table 4 in which the loadings of the previous factor are arranged along the top and down one side while their products are put in the corresponding cells of the matrix.

TABLE 4. First Factor Product Matrix

		V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8
		.33	.39	.11	.91	.53	.04	.68	.60
V_1	.33	(.11)							
V_2	.39	.13	(.15)						
V_3	.11	.04	.04	(.01)					
V_4	.91	.30	.35	.10	(.83)				
V_5	.53	.17	.21	.06	.48	(.28)			
V_6	.04	.01	.02	.00	.04	.02	(.00)		
V_7	.68	.22	.27	.07	.62	.36	.03	(.46)	
V_8	.60	.20	.23	.07	.55	.32	.02	.41	(.36)

These r's are the same as the r's in the lower left. As we do not need to add columns, they are omitted.

Now we subtract these products from the original r's systematically. This goes well till we get to the r of $V_1 \times V_7$ which equals $+0.22$ and is greater than the original r of 0.06. What does this mean? It indicates that an r of -0.16 remains to be accounted for by a subsequent factor or factors. In other words, the loadings of these two tests in the second factor, since their common possession of it causes them to be correlated negatively, are bound to be of *opposite* sign. This negative residue will occur quite frequently, in fact typically in about one half of the r's, as the residual matrix of Table 5 shows.

The fact that some variables can be positively and others negatively loaded in the same factor should occasion no conceptual difficulty. Among physical variables, for example, we might obtain a general factor of body weight and it is easy to see that body weight would influence some performances, e.g., wrestling, in a favorable way, but other variables, e.g., pole vaulting, negatively.

Can we now find the loadings in the second factor by repeating the column-adding procedure which we did for the first, aiming at obtaining the mean r of each test with all others? Any attempt to do so reveals the surprising fact that the columns now add to zero due to the balancing of positive and negative r's. Indeed, when we consider later computational checking procedures (page 161), we

shall see that this addition precisely to zero is a proof of the correctness of the preceding step. In this example, due to rounding off loadings to two decimal places, the column totals do not sum *exactly* to zero, but even with this approximation, the total of column totals is only +0.01.

TABLE 5. First Residual Matrix[a]

	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8
V_1	(.19)	.07	.24	.06	−.26	.00	−.16	−.14
V_2	.07	(.15)	−.04	.17	.15	−.06	−.23	−.19
V_3	.24	−.04	(.29)	−.06	−.42	.04	−.03	−.03
V_4	.06	.17	−.06	(.07)	.18	−.06	−.19	−.16
V_5	−.26	.15	−.42	.18	(.62)	−.08	−.10	−.08
V_6	.00	−.06	.04	−.06	−.08	(.05)	.06	.06
V_7	−.16	−.23	−.03	−.19	−.09	.06	(.34)	.31
V_8	−.14	−.19	−.03	−.16	−.08	.06	.31	(.24)
Totals	−.00	.02	−.01	.01	.02	.01	.01	.01

$T = .01$

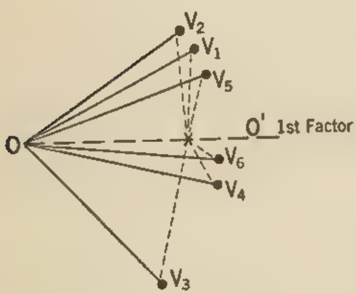
[a] The communality estimates used for next computing factor were: .30, .20, .40, .10, .50, .05, .30, and .30.

If we consider the geometrical model, this finding will not continue to surprise us. Getting the first factor is essentially running out a dimension from the origin through the *center of gravity* of the swarm of points represented by the ends of the test vectors (Diagrams 7 and 8). About this centroid they now sum to zero. There is no new direction we can take to find the second dimension; we are already at the center of the swarm. But the swarm obviously *has* a second dimension—indeed several dimensions—of spread additional to that from left to right (Diagram 8). We therefore *reflect* all the points on one side of the first axis to lie as far away on the *opposite* side as they stood on the first side. This is shown in the second part of Diagram 8 where the reflected points are shown as hollow circles[4] lying as far above the first centroid line OO' as the solid circles originally lay below it.

These points no longer balance about the first centroid. They lie

[4] The reflection takes place only after the first factor is taken out. Therefore the test vectors are not reflected about the first origin, O, but about the line OO'.

in a cluster at one side (shown in Diagram 8), and it is possible to find a new center of gravity and to run the axis of a second factor through it. The spread of the points about this new center of gravity from the origin formed by the first centroid is less than in the first factor. Indeed, each succeeding reflection requires less movement to track down the new centroid, and each new centroid has less scatter of points about it, i.e., each move reduces the amount of individuality of position unaccounted for. The points are gradually being tracked down by these successive moves and shepherded into a restricted area as are sheep by a well-trained sheep dog.



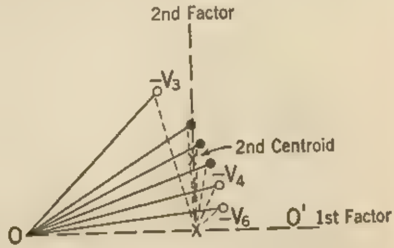


DIAGRAM 8. Reflection of Vectors with Respect to the First Centroid.

The algebraic process of reflection[5] requires that we alter the sign on the test number itself (at the same time appropriately altering the naming and meaning of the symbol) and then change the sign of every one of its *r*'s with other tests. For if sociability correlates +0.5 with quickness, it is obvious that when we measure the variable from the opposite end, thus making it unsociability, it will correlate exactly the same amount but negatively, i.e., −0.5, with quickness. Now our purpose is to reflect all these tests that have mainly negative

[5] If this reflection process is not to be a taking of liberties with signs, it is obvious that everything has to be done in conformity. There are two ways of gaining conformity. We can reverse the meaning of the variable when we reverse its sign, as indicated here, or we can retain its meaning but reverse the sign of every loading obtained for it while it is in the reversed condition. The former is more immediately intelligible; the latter is so much more convenient as to be the generally adopted practice in factor analysis.

correlations in the residual matrix (Table 5). For as Diagram 8 shows, we need to bring all the test vectors on to *one* side—the positive side—of the old center of gravity. When they are all positive, we shall then find the new centroid.

REFLECTION IN THE RESIDUAL MATRIX

Let us see which tests to reflect. On looking at Table 5, one sees that in the first residual columns tests 3 and 5 would have the highest total if it were not for the communalities of these tests. Since the communality will not change its sign when we reflect the other r's (for the communality is the correlation with the pool in the direction that we *eventually* give to the whole pool), the reversal of 3 and 5, now so largely negative, will give more positiveness to the pool than any other reversals. This business of reflecting, however, can become as exasperating as trying to hold three footballs in two hands; for as we make r's positive as a whole for one variable, we make some individual r's in the column negative for other tests. For example, in reflecting 3 for the benefit of the whole column, we make its correlation of +0.24 with test 1 become negative. So test 1 now needs to be reversed.

In the end it will be found that tests 1, 2, 3, and 4 (test 5 has finally to be reflected back again!) need to be reversed in order to make every column add positively even without its communality (which latter has ultimately to throw its weight in the direction of the

TABLE 6. First Residual with Positive Totals

	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8
V_1	(30)	07	24	06	26	00	16	14
V_2	07	(20)	−04	17	−15	06	23	19
V_3	24	−04	(40)	−06	42	−04	03	03
V_4	06	17	−06	(10)	−18	06	19	16
V_5	26	−15	42	−17	(50)	−08	−09	−08
V_6	00	06	−04	06	−08	(05)	06	06
V_7	16	23	03	19	−09	06	(30)	31
V_8	14	19	03	16	−08	06	31	(30)
Totals	1.23 (—)	.73 (—)	.98 (—)	.50 (—)	.60	.17	1.19	1.11

Total of column totals = 6.51; $\sqrt{T} = 2.55$; $m = 0.392$.

total). Usually just about half the tests need to be reflected in the first residual in order to make all column totals positive. Incidentally, it *may* be necessary to make reflections even in the first, original correlation matrix; but for simplicity of exposition we took a matrix, such as is commonly found among abilities, where the majority of r's are positive to begin with. The first *residual* with reflections completed is shown in Table 6. In practice, as described later, this reversal is done on the side in order to avoid all the erasures of repeatedly reflected signs!

Before adding the columns, we have to put in communalities appropriate to the amount each test has of the second and remaining factors. For it is not good practice to leave in the matrix the residuals from the communalities that were guessed in the first matrix. If an earlier communality was in error by being, say, 50% too much, and if half of the original now remains as a residual, this will be in error by being 200% too much! Consequently, it is better to make a best estimate of each communality *afresh* at each residual matrix, in the scale of the residual r's there found. However, for the computational purposes of checking the subtraction of the product matrix, it is necessary first to have in the literal communality residuals as we did in Table 5. Once this checking is accomplished, we can wipe out the residual communalities and estimate the communalities afresh. The manner of estimating these communalities was not explained, in the interests of avoiding complication, at the first matrix. One of the most widely used methods is to insert as a communality the highest r in the column for the given test. Now we recall that the communality equals the correlation a test would have with itself as the result of the common factor only. The method of communality estimation here adopted assumes that this correlation will be about the same as its highest correlation with any other test. The whole question will be taken up as a special issue in Chapter 10, but for the present we shall point out that this common method is a little rough and that a slight improvement upon it is gained by taking a value somewhat larger than the largest r in the columns with large r's and smaller than the maximum r where all r's in the column are small. This has been done in the first residual matrix above, rounding the estimates to remind us they are but approximations.

We now add up all columns as before, find T, the absolute total of the column totals, and divide each column total by $\sqrt{T}$ to get the

loadings. These are the loadings of the tests in the second factor. However, a slight complication has to be watched here. Since some tests—1, 2, 3, and 4—were reflected, the signs of their loadings in this factor will be reversed, and in this case, therefore, they will be *negative*. All later factor loadings of these tests will also have to reflect their signs, i.e., be taken as negative, at least unless and until we find occasion in some later residual to reflect the variables back again. The rationale of this has already been discussed in footnote 5, page 54—that in this, the second factor, they have a *positive* loading when measured in an opposite direction from that which is normal for the test.

THE PRODUCT MATRIX

At this point with the second factor extracted and the second residual left, the researcher is ready to repeat the whole procedure of calculating a *product matrix* (remembering that some of the loadings set around *this* matrix have negative signs). He then completes the cycle, subtracting the product matrix from the first residual matrix to get a second residual, reflecting whatever tests are necessarily reflected to make all columns positive, and adding up the columns to get the new loadings. The steps in this are left as an exercise for the student; here we shall merely record the end of each cycle, setting out the second and third residual matrices (Table 7) by which he may check his results.

It will be seen at once from the third residual that there is practically nothing left. Indeed, we may assume that apart from the sampling errors of the original r's with which we started and the slight errors introduced during the computations by rounding off calculations to two decimal places, this residual would be a clear set of zeros. A test for deciding whether the last small residues are a faint but real factor or a mere smudge of error will be discussed later (Chapter 17).

The extraction of factors is thus a series of cycles, each having the five steps of reflecting, estimating communalities, adding, dividing by $\sqrt{T}$, and computing a product matrix. Each cycle slices off a layer from the correlations, and the repetition of this process ends with the cycle that leaves no residue of correlation. In the present example, three cycles devoured the correlations and we may assume that if our communality estimates were correct, all the observed correlations

TABLE 7. Residuals at End of Two Further Cycles
Second Residual Matrix[a]

	+.48	+.29	+.38	+.20	−.24	−.07	−.47	−.44
−.48	(.07)	−.07	.06	−.04	.14	−.03	−.07	−.07
−.29	−.07	(.12)	−.15	.11	−.22	.04	.09	.06
−.38	.06	−.15	(.26)	−.14	.33	−.07	−.15	−.14
−.20	−.04	.11	−.14	(.06)	−.23	.05	.10	.07
+.24	.14	−.22	.33	−.23	(.44)	−.10	−.20	−.19
+.07	−.03	.04	−.07	.05	−.10	(.05)	.03	.03
+.47	−.07	.09	−.15	.10	−.20	.03	(.08)	.10
+.44	−.07	.06	−.14	.07	−.19	.03	.10	(.11)
Totals	−.01	−.02	.00	−.02	−.03	.00	−.02	−.03

Total before reflection = −0.13

Third Residual Matrix[b]

	+.21	−.34	+.48	−.34	+.65	−.16	−.34	−.31
−.21	(.06)	.00	−.04	−.03	.00	.00	.00	.00
+.34	.00	(.08)	−.01	−.01	.00	−.01	−.03	−.05
−.48	−.04	−.01	(.07)	−.02	.02	−.01	−.01	−.01
+.34	−.03	−.01	−.02	(.08)	.01	.00	−.02	−.04
−.65	.00	.00	.02	.01	(−.02)	.00	−.02	−.01
+.16	.00	−.01	−.01	.00	.00	(.07)	−.02	−.02
+.34	.00	−.03	−.01	−.02	−.02	−.02	(.08)	−.01
+.31	.00	−.05	−.01	−.04	−.01	−.02	−.01	(.10)
Totals	−.01	−.03	−.01	−.03	−.02	.01	−.03	−.04

Total before reflection = −0.16

[a] The communality estimates used for computing factor 3 were: .10, .20, .30, .20, .40, .10, .20, and .20, for variables 1 through 8.

[b] The figures around each residual matrix are the loadings on the previous factor from which the product matrix subtracted from the previous residual was calculated. It will be noticed that the signs of these along the top are reversed. This is merely part of a trick for easier calculation: It is easier to calculate the product matrix with reversed signs and *add* than to correct signs and subtract.

of the eight variables can be accounted for by three common factors, plus specifics.

Incidentally, the terms common factor, general factor, specific factor, and group factor, as used in the discussion here, have reference to the extent of the factor in a particular matrix of variables. If one and the same factor extends through all variables, it is called a general factor. For example, general mental capacity is a general factor in a matrix of ability measures, and mechanical aptitude may be a general factor in a battery composed entirely of mechanical aptitude tests. A common factor is a factor common to as many variables as one likes to name, and obviously it must cover at least two. By a specific or unique factor is meant a factor which appears in one test only and this will show up only by the fact that all the variance in that variable cannot be accounted for by common factors. Strictly, such a remnant is both specific factor and errors. By a group factor is meant something common to a group of variables within the matrix but not extending through all, i.e., it has zero loadings in the other members of the battery. Thus a set of verbal-intelligence subtests might have a verbal group factor in common additional to that general ability factor they share with all the other subtests in the battery comprising the matrix.

Questions and Exercises

1. Describe the line of reasoning by which Spearman developed a mathematical test for the presence of a single general factor in a set of variables and explain how this was extended by Thurstone's multifactor analysis to make a test for the presence of several general factors.
2. Suppose five variables are correlated according to the following table:

	V_1	V_2	V_3	V_4	V_5
V_1	(.80)	.87	.75	.38	.27
V_2	.87	(.60)	.36	.43	.55
V_3	.75	.36	(.50)	.52	.23
V_4	.38	.43	.52	(.50)	.37
V_5	.27	.55	.23	.37	(.40)

Using r_{13} and r_{45} in the role of reliability coefficients, calculate r_{gg} by means of formula (4) of this chapter. Note that the choice of these two correlations automatically determines the four to be used in the numerator.

3. Attempt to arrange the variables of question 2 in hierarchical order and explain why hierarchical order is a test of a single factor.
4. Would there be any meaning attached to a result obtained by the substitution of negative numbers for either or both denominator correlations in formula (4)? Do the same arguments hold for the use of negatives in the numerator? Notice especially that the denominator correlations are substitutions for reliability coefficients in formula (3).
5. By applying Spearman's tetrad difference equation, state whether the correlation table in question 1 appears to show one general factor or more than one. How many times must this equation be applied to the above table before such a result can be determined?
6. Calculate the set of first factor loadings of the variables in question 1, using methods of Chapter 3, and from them prepare a first factor product matrix as in Table 4 of this chapter.
7. Using the first residual matrix of Table 6, calculate the second factor loadings and second factor product matrix for these data. Then carry out the subtraction process to obtain the second residual matrix of Table 7.
8. Describe an elementary method of estimating the communalities and give some justification for it.
9. Starting with the second residual matrix of Table 7, reflect appropriate columns, insert a set of guessed communalities, and carry out the process of extracting the third factor. Depending on the choice of communalities, this third residual matrix should compare exactly, or closely, with that of Table 7.
10. What is the meaning of group factor, general factor, and specific factor?

CHAPTER 5

Rotation of Factors for Scientific Meaning

From the detailed computational procedures of the last chapter we may seek relief momentarily in broader prospects. We have extracted our precious factor loadings and, for tidiness sake, we may now pick them out from the bottom line of the various computing sheets of the last chapter, among which they are scattered, taking them from the bottom of the reflected residual matrices and arranging them securely in a single *factor matrix* as in Table 8, with due regard to the signs resulting from reflections of variables.

COMMUNALITY AND COMMON FACTOR SPACE

The column on the right is headed h^2, the agreed symbol for communality, and represents the amount of *all three of these common factors* which any particular test has. These figures are not obtained by adding those estimated communalities which were used to complete the residual matrix at each cycle in the factor extraction. Instead they are independently obtained from the actual factor loadings of the factor matrix, by squaring each of the three loadings and adding.

It may be recognized that this sum of squares represents the square of the length of the test vector in the common factor space. This will be clear if the student thinks of two factors only, when by Pythagoras' theorem the squares of the two projections of a vector V will equal the square of the length of that test vector. For as shown in Fig. 1, Diagram 9, where s_x and s_y are the loadings respectively on F_x and F_y the test vector is the hypotenuse of the right triangle formed with the loadings. This theorem extends to space of any number of dimensions.

It may seem surprising that the tests are now represented by vectors *of various lengths*, according to (the square roots of) their com-

munalities, though we started out in our spatial example by drawing them all of the same length. (In fact, we drew them all of *unit* length to indicate that all the test results were reduced to standard, comparable scores and variances). Actually each test is still of unit

TABLE 8.

Common factors

Test	F_1	F_2	F_3	h^2
1	.33	−.48	.21	.38
2	.39	−.29	−.34	.35
3	.11	−.38	.48	.39
4	.91	−.20	−.34	.98
5	.53	.24	−.65	.76
6	.04	.07	.16	.03
7	.68	.47	.34	.80
8	.60	.44	.31	.65

Specific factors

Test	F_{s1}	F_{s2}	F_{s3}	F_{s4}	F_{s5}	F_{s6}	F_{s7}	F_{s8}
1	*.79*							
2		*.81*						
3			*.78*					
4				*.13*				
5					*.49*			
6						*.98*		
7							*.45*	
8								*.59*

length, i.e., of variance *one, if we count in the specific factor peculiar to each test.* But the factor matrix normally has written in it *only the common factors*; for it is in these that we are most interested, as being widespread psychological, biological, or social influences. It is this omission of the variance unique to each which results in the unequal length of test vectors as illustrated by V_2 and V_8 from Table 8, drawn in Fig. 2, Diagram 9. Here there were only two common factors, but if the specific factors were written in, they would appear as eight extra factors, as written for the specifics in that example in the lower part of Table 8. When squared, and added to the communality, they would bring the figure up to unity. In fact, they are

generally obtained by taking the communality from unity and taking the square root.

The test vectors are therefore of different lengths only in the *common factor space*, and each has an extra dimension of its own tacked on to the common factor space, the projection in which restores it to

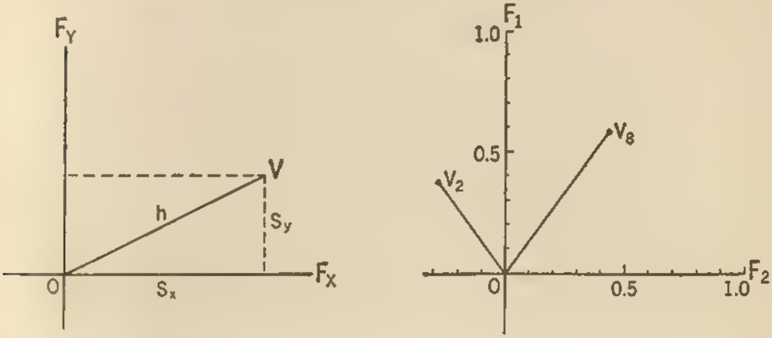


Fig. 1. General Case **Fig. 2. Current Example**

DIAGRAM 9. The Relation of Variable Projections to Communality.

unity. No other test intrudes upon this space. These private worlds of particular tests, which we call the specific factors, will be examined critically later, but for the present it is more important to understand the meaning of the common factors more fully.

TEST VECTORS

The first point to remember about the loadings in the common factors, as they emerge from computation, is that *in one respect* they are accidental. If we plot the positions of the test vectors (a vector, it will be remembered, is simply a line of given length and direction) by these given projections on the factor coordinates, we arrive at a fan of test vectors in space. We have in fact restored the structure of the test vectors as originally drawn in Diagram 5. It will be recalled that in the geometrical model approach, we arrived at the dimensions of the space necessary to accommodate the tests, from inspection of the configuration of vectors. We now see how this model can be constructed without literally beginning by drawing the correlation angles. The actual position of the test vectors in our present problem have been drawn from the calculated projections of Table 8. Though we shall not make a physical model we can represent it in Diagram 10 where the three-dimensional space is viewed from one direction at a

time, in three drawings: a front elevation, a side elevation, and a plan—as in representing *any* solid object.

But these drawings are accidental to the extent that the frame of coördinates is in an arbitrary position and can be spun around, as we realized in the second chapter. The real, unchangeable thing is the fan or configuration of test vectors. This is the solid object which has its angles fixed, but the viewing lines, like the cross wires of a microscope, can be rotated wherever we please. In short, our ground plan, as given by the factor matrix, has not started with north at the top but with whatever chance direction our computational procedure happened to give us. For different courses of the centroid analysis, e.g., chancing to reflect a different combination of variables in the sign rectifying process, and the use of different analytic methods from

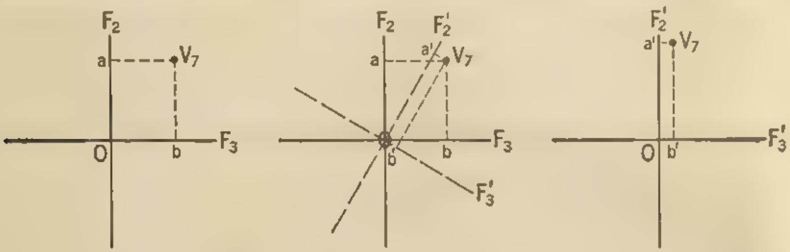


Diagram 10. Changes of Test Projections with Rotation of Axes.

the centroid, e.g., the Holzinger-Spearman bifactor process, would start us out with a different set of loadings in the three factors, though the positions of the vectors in relation to one another would remain the same. There is nothing sacred about the particular position of coördinates provided in the unrotated or initial factor extraction.

Now the statement that the test vectors themselves (or the configuration as Thurstone calls it) do not change when we rotate the coördinates implies that the projections on the various possible coördinate systems are equivalent and can be regularly transformed one into another. Thus in Diagram 10, Fig. 2, which shows both the original coordinates of Fig. 1 and those coördinates spun through 30°, from (F_2, F_3) to (F'_2, F'_3), the point V_7 finishes with a projection at a' instead of at a; but the relation of the new projection Oa' (as set out afresh in Fig. 3) to the old one Oa is a comparatively simple

function of the angle through which we have turned the coördinates and the old projections. In fact

$$Oa' = Oa \cdot \cos 30° + Ob \sin 30° \tag{9}$$

or in general, where θ is the angle of rotation and s_1 and s_2 are the two loadings:

$$\left.\begin{aligned} s_1' &= s_1 \cos \theta + s_2 \sin \theta \\ s_2' &= -s_1 \sin \theta + s_2 \cos \theta \end{aligned}\right\} \tag{10}$$

This shift would change the loadings of the tests given in Table 8 to those in Table 9. There the first factor loadings remain the same; but by swinging the coördinate system of the new drawing of Factors 1 and 2, we could change these also so that *all* the factor loadings of the tests would be different.

TABLE 9. Projections After a Shift of Factors 2 and 3

Test	F_1'	F_2'	F_3'
1	.33	−.31	.42
2	.39	−.42	−.15
3	.11	−.09	.61
4	.91	−.34	−.20
5	.53	−.12	−.69
6	.04	.14	.10
7	.68	.58	.06
8	.60	.54	.05

$\cos 30° = 0.87$

$\sin 30° = 0.50$

Although the individual projections of a test change, increasing on one axis as they decline on another, two things remain constant—the sum of the squares of these projections and the correlations among the tests as calculated.[1] The first follows again from Pythagoras' theorem, since the test vector remains the same length through the rotations and is the hypotenuse of the triangle formed with the two projections. The second can be tested empirically by applying formula (2) to Tables 8 and 9. Thus in Table 8, $r_{12} = (0.33)(0.39) + (-0.48)(-0.29) + (0.21)(-0.35) = 0.20$; and in Table 9, $r_{12} = (0.33)(0.39) + (-0.31)(-0.42) + (0.42)(-0.15) = 0.20$. This may be compared, incidentally, with the value 0.20 in Table 1, the correlation matrix.

[1] The latter is equivalent to saying that the scalar products of the test vector projections remain constant.

Without at the moment pursuing further the mathematics of the transformations occasioned by rotation, we can see that we do not lose anything in rotating, for the new information is always equivalent to and changeable into the old, while the communalities of the test loadings remain constant for any position of rotation. Indeed, the mathematical properties of the various rotations are so equivalent that a mathematician is inclined to say that any one position is as good as any other. And since there are an infinite number of points of the compass at which the coördinates can come to rest, he is inclined to say, "Let us accept the position given initially by the computations and save ourselves further work."

We have mentioned, however, that different computational systems start us off at different positions. The particular computational position therefore has no sense except as an immediate convenience and, as will be seen shortly, there is *one* rotation position to be found which makes especially good sense. Consequently after extracting factors it is always necessary to enter on a process of rotation to find this most meaningful position.

The psychologist or social scientist cannot be content with factors that are merely mathematical conveniences. An analogous situation is that of systems of latitude and longitude where a grid for measuring in two directions at right angles could be started anywhere. In longitude, it is true, we begin at a point of merely historical importance—Greenwich; but latitude is fixed in regard to real features—the poles and the equator. In the case of factor analysis, the scientist wants each factor to correspond to some unitary influence with which he is familiar on *other* and general scientific grounds—some influence which he has reason to believe is a functional unity in nature. Consequently, he argues that there is *one* position in the rotation which corresponds to the real factors and that all the other positions encountered are mathematical transformations of this real position—false claimants which we have not yet succeeded in eliminating. Like a a man in a hall of many mirrors we see seemingly countless images of the same object, all behaving in the same way, and for the moment we are perplexed about deciding which are merely reflections and which is the object. It is the weakness of our mode of computation that it gives the true factor position and all its shadow equivalents in a single system. We therefore have to apply a second method of examination to pick out the rotation which corresponds to reality.

SIMPLE STRUCTURE

Without venturing to enter at this stage upon arguments as to why certain rotation positions give greater meaning to factors and greater usefulness than others (see Chapters 12 and 13), we shall adopt as our standard here the goal of simple structure. *Simple structure,* as propounded by Thurstone, is indeed the most widely used and widely practicable criterion for finding a uniquely meaningful position. Thurstone bases this notion on the broad scientific principle of parsimony which Newton, among others, expressed in the belief "Natura est simplex." According to this axiom if we have several alternative hypotheses, each fitting equally the given facts, we should decide among them by taking that which is the simplest, i.e., that which requires fewest conditions and least bolstering by supplementary hypotheses.

In terms of factor analysis, Thurstone argued, this means that any one test should have the simplest possible factor constitution and reciprocally, the estimation of any one factor should require the combination of only *some* of all the possible tests. This means in terms of the factor matrix that every test should have some zeros in its row, i.e., that some factors should not load it, and that every factor should have some zeros in its column, i.e., that not all tests should be affected by it.

In a factor analytic solution rotated to simple structure there is actually a double application of the simplicity or parsimony principle. First we have represented many variables by a few common factors and secondly we have distributed these factors to give the simplest explanation for that number of factors.

Nevertheless proper application of parsimony in the realm of correlation is not easy to decide, and some awkward questions that may occur to the psychologist at this point will have to have their discussion postponed to Chapter 12. Simple structure, however, need not be stated *precisely* in these particular terms of Thurstone; and it is clearer, perhaps, to state it in slightly different terms, namely, that we should *not* expect any one psychological influence to have appreciable effects on more than a fraction of the total personality. Thus intelligence may affect a person's abilities in various directions, his powers of memorizing and his levels of information; but there is no special reason for it to affect significantly the extent of his sociability

or his irritability or his proneness to catch colds. Similarly in the social realm, an economic factor such as a nation's standard of living may affect birth and death rates, expenditures on education, etc.; but it is unlikely to be related to the annual mean temperature, the size of the capital city, or the ratio of artistic to scientific creativity.

THE HYPERPLANE

In short, if our battery of variables contains a good assortment of measures (see the personality sphere concept, page 331), we should expect each factor to be involved in only a fraction of them. In graphical terms this means that the points representing the ends of the various test vectors should lie partly well out on the factor axis, i.e., with

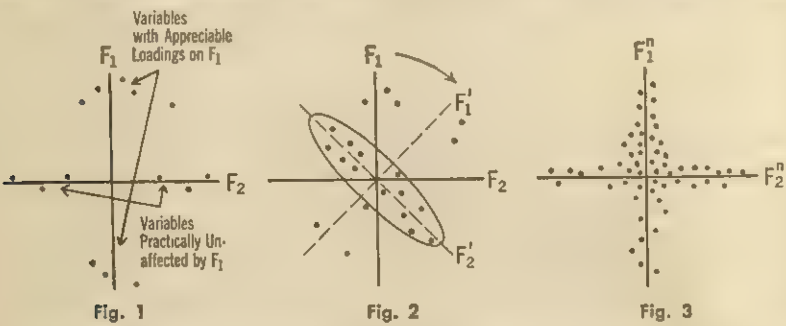


DIAGRAM 11. Simple Structure and a Clear Hyperplane for Single Factors.

appreciable projections (loadings) on it, and partly in a line through the origin at right angles to the factor, so that they have zero (or near-zero) projections on the factor, as the horizontal row of points in Fig. 1, Diagram 11, has with relation to F_1.

The line of points which has zero projections is of course not really a line. Except where there are only two factors in the problem, it is a circular disk (or a disk in multiple space, namely a sphere) which is cut by our drawing so as to be seen edgewise and thus appears as a line. If our objective in simple structure is to find a position for the axis of a factor such that a good number of points will have no projection on it, we are really searching for a disk to put at right angles to the axis. When we have found this disk (running, of course, through the origin), we have fixed the position of the factor, for it will run through the disk as an axle runs through a wheel.

By this means, therefore, *the poles of a factor have their position discovered through finding an equator.* This process is illustrated in Fig. 2, Diagram 11, where the first plot on F_1 and F_2 showed a conspicuous row of points running in a NW-SE direction clearly indicating that F_1 needed to be rotated to the right, to F_1'. Sometimes one may encounter this disk in a somewhat elliptical shape (as drawn in Fig. 2) instead of as a simple straight line (as in Fig. 1), because the rotation on F_3 (sticking up out of the paper) is not yet right, and the F_1, F_2 drawing does not therefore cut perpendicularly across the disk. Because the disk or plane is usually in a space of several dimensions—at least in the hyperspace beyond three dimensions—it is usual to call it the *hyperplane* of the given factor. It will be noticed that so long as factors remain at right angles, the cutting of the hyperplane by the plane of our two-dimensional drawings produces a line of points for the hyperplane of factor 1 which lie along the axis of factor 2, and vice versa. Hence, when simple structure is found, we usually get an effect as in Fig. 3, Diagram 11, which is taken from an actual research. However, as will be seen later (page 117), it is not always desirable to keep factors orthogonal (i.e., at right angles) and in that case the cut of the hyperplane of factor 1 will *not* coincide with the axis of factor 2 or whatever factor is paired with factor 1 in the drawing.

VALIDITY OF SIMPLE STRUCTURE

The question as to whether simple structure provides a real and legitimate criterion for determining the unique, scientifically meaningful rotation position from among the infinite number mathematically possible must be answered by factual experience. Is it in fact *possible* to find these disks of unmistakably greater density when one plots the positions of points according to the initial unrotated factor loadings obtained by the process of factor extraction? The answer from some hundreds of adequate researches in abilities, in personality ratings, and in social and physiological data is that generally the chaos of points as initially plotted will be found, after some preliminary groping rotations, to reveal the well-marked swarms of points, clear to the eye as nebulae among the stars, which constitute the hyperplanes of factors.

Correlation coefficients taken at random and thrown into a matrix may yield factors, but the plots will not yield hyperplanes. The hyperplane shows the tracks of real, organic influences in the mass of cor-

relations concerned. Occasionally, even where distinct influences are known to be operative, no structure can be found to exist because errors have made the hyperplanes too fuzzy, and conversely, instances exist where an accidental heaping of points has falsely given the impression of a hyperplane; but techniques and criteria will be discussed later to handle such situations.

The process of rotation for simple structure can be illustrated by the data of our three-factor problem, which, incidentally, is known to contain a simple structure. The initial plotting of the eight test points

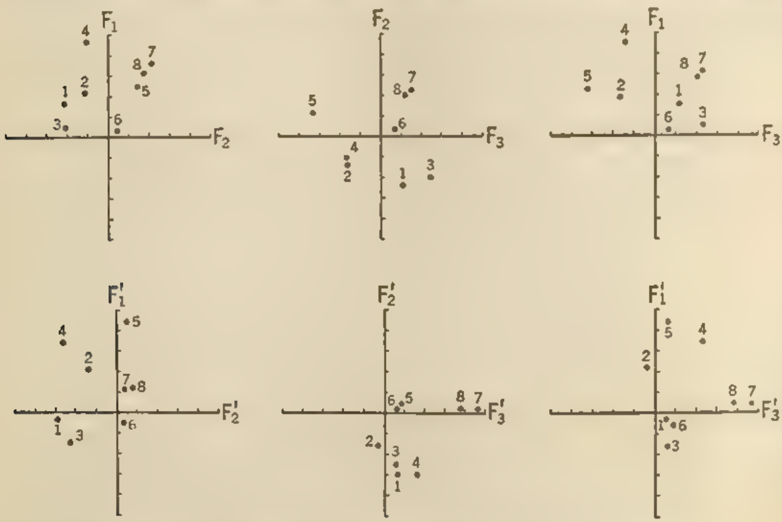


DIAGRAM 12. Shifts in the First Rotation of Three Axes.

from the original unrotated factor matrix given in Table 8 above is shown in the upper row of graphs in Diagram 12. These views show, however, no sign of structure—the points seem as randomly distributed as eight points can be. After three rotations, the detailed computations for which will be described in Chapter 12, a position was reached as shown by the three views (each looking the length of one axis) in the lower row of Diagram 12.

Though one cannot expect swarms with only eight points, it will be seen that for each factor about half of the variables are now in the corresponding hyperplanes. That is to say, some four of the points—a different four in each case—have a practically zero projection on

the factor drawn vertically (F_1', F_2', and F_3') in the three drawings. Incidentally, we should not expect them to fall *exactly* on the zero projection line, for owing to sampling errors in the original r's, we shall suppose that any loading falling within $+0.1$ and -0.1 is essentially zero. Since further rotation fails to improve on the above, it is concluded that the best simple structure is obtained and the projection of the variables on these axes are accordingly copied down in a new factor matrix which we call the *rotated factor matrix*, as follows:

TABLE 10. Orthogonally Rotated Factor Matrix

Test	F_1'	F_2'	F_3'	h_2
1	−.05	−.61	.10	.38
2	.46	−.37	−.04	.35
3	−.38	−.47	.13	.39
4	.75	−.54	.38	.98
5	.85	.08	.17	.76
6	−.11	.01	.14	.03
7	.12	.02	.89	.80
8	.10	.04	.80	.65

It will be seen that though the loadings of each and every test are different from those in the *unrotated matrix* of Table 8, the communalities (h^2 column) are the same (except for rounding errors) and, as the reader can test for himself, the correlations of any two tests when restored by calculating the cross products of the loadings will be found to be the same. For example, the correlation of test 1 with test 2 is $(-0.05)(0.46)+(-0.61)(-0.37)+(0.10)(-0.04)=+0.20$, which compares with $+0.19$ similarly obtained from the unrotated factor matrix (Table 8, page 62), from the intermediate rotated matrix, and the original value of $+0.20$ from the correlation matrix (Table 1, page 41). Thus it is shown that the rotation has not altered the angles of the test vectors among themselves or their length in the common factor space.

Before turning to further details of the actual computational processes in such rotations, we shall continue in the next chapter with the main argument in order to see the whole process in perspective before concentrating on technicalities.

Questions and Exercises

1. Explain the geometric concept of the communality of a test. If a test had a communality near 1.00, would one be justified in expecting this test to measure some quality that the other tests in its group did not measure? Why?
2. Find the communality of a test with two factor loadings of +0.75 and −0.83. assuming these to be the only factors in common with other tests of the group. Draw to scale a vector representing this communality with respect to the two orthogonal factor axes. (Note that the length will in general exceed the value of the communality, since the square root of a number less than 1 is larger than the number itself.)
3. Find the new coördinates of the end point of the communality vector in question 2 after both axes have been rotated 60° counterclockwise. Calculate the communality of the vector in terms of these new coördinates and compare with the result obtained in question 2. If only one axis had been rotated, should one expect this communality to remain constant? (Give geometric reasons for your answer.)
4. Give a trigonometric or algebraic proof for the last answer of question 3.
5. State the aims of a rotation of a factor pattern to its simple structure: i.e., how would one be able to recognize a simple structure by the factor loadings after rotations and by their graphs?
6. Describe the geometrical significance of a hyperplane, particularly with respect to its factor axis.
7. Assuming that a loading between −0.10 and +0.10 indicates a point in the hyperplane of the factor with this loading, how many points lie in the hyperplanes of factors F_1, F_2, and F_3 as they are given in Table 8? After rotation to the positions indicated in Table 10, how many points lie in each hyperplane?
8. Examine the drawings on the lower row of Diagram 12 to see whether you can find any rotation that would produce an improvement in simple structure. Measure the angle of the possible (but not better) shift on the first drawing and try to work out for yourself the general rule for the change in magnitude of projection of a point on a factor when the factor axis is shifted through a given angle.

CHAPTER 6

Factor Estimation and the Specification Equation

Granted that we have taken out the number of factors required to account for the correlations and have rotated them to a special position which gives them the greatest meaning and usefulness, what do we now have and how can we apply the knowledge? First we have the new factors, the nature of which we shall hope soon to recognize and which we need to be able to estimate in order to give every individual his new apparel of factor measurements to replace the motley rags of his original numerous scores on test variables. Second, we need to know how to use these factor measurements for various kinds of predictions by means of the specification equation. For example, from forty clinical and general tests of various kinds we might set out to assign to individuals a score in factor *A* or schizoidness of personality, and factor *B* or general mental capacity, from which to predict his performance in some occupation or his probable response to a course of psychotherapy.

ITERATION FOR EXACT VALUES

Before starting out on such further steps, it behooves us as statisticians to recognize that our foundation contains certain approximations and unreliabilities. About the unreliabilities of the original *r*'s nothing needs to be said to a reader sophisticated enough to read this book, but he should be reminded that when starting the factor extraction, we were compelled to guess the communalities, i.e., to estimate the values in the diagonals of the correlation matrix, according to principles to be explained when we come to details of computation procedure (Chapter 10). For the present it is necessary only to notice that the actual communalities with which we finish, as in Table 8 or Table 10, are not quite the same as those we estimated. The success of the

shrewdest guess is in these matters always only partial, and the error occasioned spreads itself thinly through the whole factor matrix including the final communalities.

If we wish to reduce this slight error, it can be done by repeating the factor extraction beginning with the more accurate communalities obtained at the end of the factorization (Table 8) in place of the initial guesses. For the final communalities are better than the original guesses, since they are largely derived from the actual correlations. But even this will not give a perfectly correct result and, indeed, we could repeat the whole process of successive approximation (or iteration, as the mathematician would call it) several times to converge with increasing accuracy on the true values.

In actual fact this iteration is practically never carried out, social scientists being limited in funds, time, and patience; and, at the present exploratory stages of research in most fields, any refined acuracy of iteration is not really appropriate. However, in studies that are carried out with fewer than about ten variables, iteration is desirable for even a *reasonable* degree of accuracy. On the other hand, with correlation matrices having more than ten variables in the column, and especially when we have forty or more, it is obvious that any error in guessing the communality contributes an extremely small percentage to the column total (relative to the total of so many correlations) and does not normally justify any repetition of the lengthy process of factor extraction.

Let us suppose therefore that we are tolerably satisfied with the accuracy of the unrotated factor matrix of Table 8 and the matrix rotated to simple structure in Table 10. What is our next step in using these results? Factor recognition, factor estimation, and the specification equation have been mentioned as primary objectives; but it should be added that these are usable for broadly two different aims. First, there is the scientific aim of discovering the nature of the factors at work in the given phenomena; and second, there is the more practical aim of providing an instrument, an equation, for predicting happenings to particular people in specific situations. Let us consider these in order.

NATURE OF A FACTOR

The nature of a factor is initially approached by reasoning (and sometimes by intuition or hunch) based on inspection of the factor matrix to see which variables are highly loaded in the factor and

which have nothing to do with it. Thus in Table 10 (page 71) we find factor 1 is loaded highly (0.85) with respect to test 5 and less markedly with test 4, test 2 is moderately loaded, and tests 1, 6, 7, and 8 are virtually unaffected by the factor. Suppose that we find test 5 to be a rating of carefree cheerfulness; test 4, a measure of sociability with the opposite sex; test 2, a measure of talkativeness; and test 3 (with a negative loading of -0.38), a physiological measure of anxiety. This pattern suggests the well-known temperament dimension of surgency vs. desurgency; and when we find that variables 1, 6, 7, and 8 are such as are supposed to have no relation to this pattern, our hunch is strengthened. The factor is one (measured in its positive direction) of temperamental insusceptibility to anxiety, while in its negative direction it runs to anxiety and (we may later find) to depression.

Factor 3 has high positive loadings in 7 and 8 and very little else. Suppose we find that tests 7 and 8 turn out to be respectively an analogies test and a classifications test. Evidently we are dealing here with an ability rather than a temperament trait—an ability to perceive relations and educe correlates such as Spearman hypothesized as the nature of intelligence. This identification with *general mental* capacity is strengthened when we notice that variables 1, 2, 3, and 6 with practically zero loadings are tests whose obvious nature is quite remote from ability of any kind.

Factor 2 has no really high loadings. It is characterized most by test 4 with a negative loading -0.54, and test 1 with a negative loading -0.61. These are measures, respectively, of sociability with the opposite sex and interference with a performance score through sexual stimuli. Tests 2 and 3 with slight loadings turn out to be talkativeness and anxiety. The nature of this factor is more obscure, though a shrewd clinician might hypothesize that it is strength of the sex drive.

The search for common characteristics in the loaded variables which would give a first hunch as to the nature of the factor is beset by difficulties when the loadings are not very high, and always presents possibilities of being misleading. To take a trivial, not to say frivolous, example, if two drunken men and two sober men constituted our population and one of the former had had Scotch and soda while the other had had Bourbon and soda, but the sober men had had nothing, we should obtain correlations suggesting a cluster or factor in which drunkenness and soda would be most strongly loaded. Only a person who knew that the variables—Bourbon and Scotch—contained the

common influence alcohol would recognize the role of alcohol in the drunkenness syndrome; and only the choice of a sufficiently varied population to include persons who had drunk soda but not alcohol would reduce the soda variable to its proper negligible loading in the drunkenness factor. The interpretation of factors, as we shall see later, requires such alertness to odd population conditions i.e., to effects of sampling. But primarily it requires further *experiment* with new variables to see whether the hunch suggested by the first factorization was good.

Thus before the tentative inferences made about the nature of the above three factors (one a temperamental pattern, one an ability, and one a dynamic trait) can be confirmed, it is necessary especially in the third instance to make up a test or choose an observation which by our hypothesis would most directly measure the trait (factor) supposed to be operative. We should then return to experiment with these new variables along with the old, and upon factorization we ought to find that these test variables are *more highly loaded in the factor concerned than any others.* In fact, if we are able to choose a test which we *know* is a pure measure of the hypothesized entity or function, its loading in the factor should be as near to unity as experimental error will permit.

SPECIFICATION EQUATION

Until this last step is achieved, the identification of a factor is never complete. Its nature is in some degree known, in so far as we are able to imagine some power, or cause, or common character behind all the measures which it is shown to load moderately. It is known by just the same act of logical abstraction as we use when we speak clinically of neuroticism behind all the particular symptoms of neurosis or when in economics we speak of the business cycle behind particular trends, e.g., in interest rates and unemployment, no one of which is perfectly correlated with the entity we abstract from them. During decades in which research is unable to gain much beyond such identification by relatively moderate loadings and is unable to find the perfectly saturated variable which will indisputably declare the nature of the factor, science has at least identified the factor to the extent of being able to recognize it by its loading pattern, to describe the best marker variables for it and thus to introduce it into various experi-

ments and explore its relations to other things. We are, at that stage, in the position of the biologist who has captured a specimen and preserved it though still not knowing much about it, or of the physician who has recognized a new disease by a recurrent symptom complex but still does not know its cause.

The second use of factors, in connection with prediction of individual behavior or the behavior of groups and other organisms, arises from application of what is called the specification equation. The specification equation is simply a row taken from the factor matrix (without the communality column, of course), such as the row for test 1 from Table 10, written as follows:

$$P_{v_1} = -0.05F_1 - 0.61F_2 + 0.10F_3 \tag{11}$$

This means that the variation in score in test 1 is contributed to by variations in each of the three factors to the degrees shown by the loadings. It will be remembered, however, that there is also a factor peculiar to performance in variable one which is not a common factor and which we have called a specific. This specific must be included in the equation for complete prediction though its amount can be determined only by subtracting the variance in all the common factors from unity, and thus in a sense it is a measure of our ignorance. As calculated in Table 8, the loading of the specific factor in variable one is 0.79, and for completeness we should write the above specification equation thus:

$$P_{v_1} = -0.05F_1 - 0.61F_2 + 0.10F_3 + 0.79F_{v_1} \tag{12}$$

We may now generalize this specification equation for all future references as follows, letting i stand for a particular individual and j for the particular conditions defining the performance P.

$$P_{ij} = s_{j1}T_{1i} + s_{j2}T_{2i} + \ldots + s_{jn}T_{ni} + \ldots\ldots\ldots s_jT_{ji} \tag{13}$$

Here s has been chosen as the more general symbol of what the statistician calls loadings, because to the psychologist or social scientist s means a situational index, i.e., the extent to which the particular situation j in which the performance P is measured involves (stimulates, tries out) the factor in question. Similarly T has been substituted for the mathematician's F, because, in psychological terms, we are dealing with a *source trait* (22) or, in socioanthropological data, a culture trait. In short, a factor is a general attribute, dimension,

pattern, trait (hence T above), or characteristic of an organism in the widest senses of organism and attribute.

It will be observed that the s has subscripts j and a factor number n indicating that the value of this situational index is peculiar both to the situation j and the factor in question. The factor also has two subscripts: an index number n, indicating the identity and nature of the factor, which number it shares with the situational index; and a letter i, indicating that it belongs to a particular individual and that the quantity we shall enter there represents the extent to which the particular individual i possesses that factor. The last or specific factor has j instead of an index number because it is unique to the situation j, and such situations are too multitudinous for any simple numerical indexing.

When factor analytic research has given us the *meaning* of a situation (for personality or group response) by providing the above situational indices or loadings as from our factor matrix, it becomes possible to calculate the individual's performance P_{ij} (in standard scores) if we have his factor endowments (in standard scores). Thus if a person were just average in the surgency (sociability-cheerfulness) factor above, a standard deviation below average in F_2 and half a standard deviation above in the general ability factor, his personality pattern could be written: $F_1 = 0$; $F_2 = -1.0$; and $F_3 = +0.05$. His personality could, incidentally, as far as these three factors are concerned, be represented not only by a profile of the above measures but alternatively as a point (position) in space, fixed by these projections on three axes. Inserting these values in the specification equation for variable one and giving subject i an average score on the specific factor i.e. $F_{v_1} = 0$, we obtain (see page 77):

$$P_{v_1} = (-0.05)(0) - (0.61)(-1.0) + (0.10)(0.5) + (0.79)(0) = 0.62. \quad (14)$$

That is to say, we should expect him to be 0.62 standard deviation above average in the performance P_{v_1}, which is degree of interruption of a task by irrelevant stimuli.

FACTOR ESTIMATION

It is important to stress that this is an *estimate*, because it is open to certain known probabilities of error. Thus it is like any regression equation in statistics, the error of estimate being primarily due to the

fact that we are unable directly to measure the contribution due to the specific factor T_j, but also to the error of estimate involved when we have to add performance on a set of variables to obtain the value of each of the other factors going into the equation. True, we do have an estimate also for the specific or unique factor T_j, but it is obtained only from our knowledge of the person's performance in this particular situation! We are therefore arguing in a circle as far as T_{ji} is concerned, whereas the other common factors can be estimated from the person's performance in quite different, independent situations. In fact if we were going to use only the variables *in this particular matrix* to estimate these factors and then use the factors to estimate variables, the procedure would be pointless; we might better take the direct measures of the given individuals on the given variables! But once the individual's factor endowments are estimated they can in fact be used in innumerable specification equations to predict a great variety of *new* performances.

Let us next look more closely at the manner in which any given factor is assessed for any given individual. If we knew the perfectly loaded test in any factor, we could use that single test to measure an individual's endowment; but generally the best loadings we can get are somewhere in the region of 0.5 to 0.9. Such tests have other common and specific factors (for theoretically there is no reason why the specific should not be *several* unique factors) in them. So we need to combine several distinct tests in order to estimate the factor in question. Adding them up, the common elements will accumulate and the specifics will become less important and nullify one another.

Our first step, however, should be to rule out from the group of reasonably highly loaded tests for this one factor any that also *share any other common factors*. For example, in measures of the numerical ability factor we should not have *more than one* test that also has appreciable loadings in the spatial ability factor, no matter how good the test to be added may be in the numerical factor. (Of course, it cannot be *perfectly* good in the numerical ability factor, loading it +1.00 and still have any spatial factor, since the total communality is but one.) But generally we have to be content with tests having loadings of about 0.6; and indeed five to a dozen of such tests can give a highly reliable estimate of a factor providing they are only bringing in mutually devaluing specifics besides. For, as indicated above, if

they also bring in two or three measurements on some foreign common factor, the estimate on the first factor will become *systematically* biased.

WEIGHTING

Among the set of half dozen or so satisfactory tests for achieving the estimate obtained by going down the column for the factor (Table 10) and picking out the high loadings, dropping those with high loadings in common factors elsewhere in the factor matrix, we might think to weight those most highly which have the highest loadings. It is certainly necessary to weight those tests negatively which have negative loadings, for the negative loading sign means that a high score on this test should reduce the individual's score on the factor. But more detailed weighting than this is not much used in practice because the true weights by which the standard scores in the various tests have to be multiplied before being combined are not the factor loadings themselves and require further calculation disproportionate to the gain in accuracy.

Any student familiar with partial and multiple correlation will realize that when tests correlate with a criterion but also correlate among themselves, some allowance has to be made for these interrelations when using the test correlations with the criterion as a basis for getting a best weighted combination in estimating the criterion. That is the situation here, with the factor in the role of criterion. For methods of obtaining these weights accurately from the loadings and the r's between variables the student is referred to standard statistical textbooks on computing beta weights in the multiple correlation coefficient by such methods as those of pivotal condensation, the Doolittle method, etc. (**45, 46, 95, 120, 126**).

Actually, accuracy gained from using these weightings is frequently not worth the trouble of all the computation involved. For many studies, if the picked group of variables has reasonably high correlations with the criterion, it suffices simply to add their standard scores with equal weight. For example, such subtests as analogies, classifications, series, and inferences, of equal length, usually have approximately equal loadings in the general intelligence factor, so that in estimating endowment in the latter from a test we commonly add the different subtests unchanged. (Indeed, we also omit to convert them first to standard scores!)

PREDICTING THE VALUE FOR A VARIABLE

When we have the individual's personality dimensions in terms of the above source traits estimated with suitable accuracy from the scores on the variables and also have the meaning of the particular performance situation in terms of a set of situational indexes, the performance in the given situation can be predicted by the above specification equation. As a fuller illustration of this than has been provided by the single specification equation worked out above, let us take the same factor matrix and compare estimates on two individuals in the same situation and the same individual in two different situations. Let us suppose the individuals A and B are found from their factor estimates, to have the following endowments, at the time, in the four dimensions (in standard scores).

	T_1	T_2	T_3	T_4
Person A	.8	−.4	2.3	−.1
Person B	1.2	1.0	−.7	.2

If we accept the interpretation of these factors just attempted above, the first person will obviously be verbally described as a person of somewhat low intelligence and strong sex drive. Individual B will be outstanding for his surgent temperament and will have rather high intelligence and an average sexual disposition. In the performance of variable 4—degree of sociable activity with the opposite sex—we shall come out with the following after substituting in the specification equation:

$$\begin{aligned} P_A &= (0.75)(0.8) + (-0.54)(0.4) + (0.38)(2.3) + (0.13)(-0.1) \\ P_B &= (0.75)(1.2) + (-0.54)(1.0) + (0.38)(-0.7) + (0.13)(0.2) \end{aligned} \qquad (15)$$

which equal 1.24 and 0.42, respectively, for A and B, indicating that person A is decidedly more active in this situation.

In the above two equations the situational indexes are the same while the endowments are different; but when we predict for one person with respect to two situations, the endowments remain constant and the indexes alter, as when we ask whether A will be more outstanding in his talkativeness or in his performance on a classification test and obtain the following answer:

$$
\begin{aligned}
P_{A2} &= (0.46)(0.8)+(-0.37)(-0.4)+ \\
&\qquad (-0.04)(2.3)+(0.81)(-0.1)=0.34 \\
P_{A8} &= (0.10)(0.8)+(0.04)(-0.4)+ \\
&\qquad (0.80)(2.3)+(0.59)(-0.1)=1.84.
\end{aligned}
\tag{16}
$$

Two criticisms of the use of the specification equation need to be inspected at this point. First, it is sometimes alleged that since we use the test scores to estimate the factors and then use the factors to predict the test performances, we are merely moving in a circle, as mentioned two pages back. Secondly, and in connection with this, it is said that the factors are unnecessary intermediate variables in predicting from a test performance to a criterion, e.g., success in a particular occupation.

In discussing the first of these possibilities a couple of pages back we saw that the charge of pointlessness would be partially true if we included, in estimating the factors, figures from that very row of the factor matrix which contains the variable to be estimated and which sets out the specification equation eventually to be used in the estimation. Usually it is as unnecessary as it is undesirable to drag in this particular variable; there are in a large matrix plenty of other tests from which to estimate the factors. Moreover, as research progresses, we increasingly depend on special, *ad hoc* tests from basic research to measure the factor. The immediate applied research which gives the particular specification equation to be used, showing for example, how much intelligence needs to be weighted in the prediction of a particular performance, is not used at all in *estimating* intelligence. We have our standard tests for this. As indicated above, the only factor that has to be estimated in circular fashion is the *specific factor* which is generally, therefore, regarded as so much error of estimate.

WHY FACTOR?

Far wider issues have to be raised in meeting the second criticism. If an industrial psychologist obtains an r of, say, 0.6, between a questionnaire and some rating of success in an occupation, he is inclined to use the test without further ado and to be unsympathetic to demands that he factorize (or obtain understanding by some other means) as to *why* the test predicts. Dispensing with intermediate variables, i.e., omitting estimation of the factor from tests and of the occupational

performance from factors, is, however, at best a short-sighted economy and at worst an unenlightened resistance to scientific curiosity and scientific method themselves. For all scientific inquiry works by introducing hypothetical constructs, as the philosophers might call them, between one set of facts and another. It attempts to explain why one set of facts, e.g., the dependent variables, behave in a certain way when another set, the independent variables, behave in the way observed. And such understanding gives greater control everywhere.

For the desire to understand why and through what factors the given tests manage to predict performance is not only a need to satisfy scientific curiosity. As with all scientific understanding, factor analysis yields dividends in the form of gains in practical control. For in fact it has often happened that on trying out his particular patent battery of tests with another sample of employees, the industrial psychologist has found that their predictive value has melted away—the original correlation no longer appears. Or to take an example from early clinical experience, after the clinician had used form boards as a measure of intelligence with very young children, he was grievously led astray by the same device as an intelligence measure for older persons. It was not realized that though the factor of general ability is highly loaded in form boards when using a population of children it is not so loaded in adults. The factor of intelligence manifests itself differently at the different mental age levels. Unless we know *why* a given correlation appears in some applied problems, i.e., unless we know the scientific entities that are operative, we are likely to be led astray very often.

When the psychologist uses factor analysis to establish source traits, he proceeds to find out something about the natural history of these source traits. Thus, after Spearman had shown in 1904 the existence of a general ability factor, the ensuing years of research on general mental capacity demonstrated that this factor normally increases only up to 14 years of age, that it is largely unaffected by environment and physical condition, that it affects success in spatial, verbal, and numerical thinking and that it is partly responsible for goodness of memorizing. The fraction which we call the I.Q. was found to be relatively constant, and so on.

Knowledge of the natural history of the factor, i.e., of the general psychological laws that apply to it, is therefore essential to understanding the meaning of psychological tests and knowing what to

do about their indications in a given case. A test which merely yields a correlation with the criterion may be a composite measure of two or more factors, and with the passage of time one of these may behave in a very different fashion from the other. It is such variations of the operation of a mere test with circumstances, sample, age, etc. that make an insightful use of factors, via the specification equation, scientifically and practically superior to blind empiricism, even though there is some slight loss of accuracy through introducing the computations necessary to estimate the intermediate variables. Quite apart from this fact that most tests are measures of a whole mixture of factors there is the further economic objection that the task of establishing the requisite age development, practice development, and prediction values for many thousands of specific tests is far greater than doing so for a more limited number of important factors.

The preceding paragraph should also be sufficient reply to the uninformed criticism sometimes made of the factorial conception to the effect that factors are static, or that they deal only with constitutional or nondynamic traits. A factor or source trait may be innate in origin or it may be a pattern of training imposed by a social institution; it may be dynamic in nature or it may be an ability; it may fluctuate from day to day or it may remain remarkably constant. After the factor has been established, these further facts of its natural history normally become known and are taken into account in making estimates for prediction in the specification equation, especially in regard to change with time, place, and stimuli. But even without such advances in the psychological knowledge of personality traits or culture traits, the purely mathematical use of the specification equation remains free, as indicated above, of many objections that apply to the individual test. In this case, however, correct usage requires that the estimates of the factors be made and used at the time when the performance is to be estimated.

Another cogent argument for the use of factors is an extension of the above argument of economy. Our measuring should remind us that it is the same personality with the same factors in it that enters classroom, industry, army, or clinic. Yet it is a reflection on the planning of research in industry, education, and clinical psychology that each has expended effort on the accumulation of particular local "patent medicine" tests, each considered satisfactory because it has some moderate correlations with the criteria in the particular field, though for no known reason. For example, a straight score on

a certain biographical inventory is found to give fair predictions of infantry officer success, while various mechanical aptitude tests have given correlations with success in some factory operations, and the Rorschach test gives correlations of 0.3 to 0.4 with clinical evaluations of patients. None of these has been designed and constructed after prior research on real personality factors. Each has been built up in one particular field of applied psychology and though it claims to assess the essential personality, it would not be considered suitable for that purpose by a worker in any other field. So widely standardized and investigated a test as the Strong Interest Blank is used by industrial, vocational selection workers but would not be considered of any relevance by clinicians, and so on. It is not the slight differences of testing conditions in classrooms, factory, army, or clinic which have accounted for or required this duplication, but the absence of scientific organization to center research on the real structure of personality instead of on single tests and quick predictions from inadequate research for particular purposes.

Similar advantage could doubtless be cited in the social sciences from concentrating research upon some clearly formulated factors of universal validity instead of upon *a priori* indexes, e.g., of business activity, which may or may not turn out to be very central to any important prediction. A fact of economics is also a fact of sociology, cultural anthropology, and social psychology, and can be understood only when structured by factorization of the whole.

PARTICULAR VARIABLES OR BASIC CONCEPTS?

Gains from shifting research effort from particular variables to basic concepts in terms of factors thus extend beyond the realm of pure scientific understanding into questions of efficiency of organization of applied practice. When a dozen or so primary personality factors and abilities have once been measured, they can be used in all fields. It happens that at present the established natural history about those factors is relatively thin. We do not yet know the nature-nurture ratio for most personality factors, nor what function fluctuation[1] they undergo with circumstances. Consequently the rewards in

[1] Function fluctuation refers to that nontrend change in a measurement which is due to real changes in its strength and not to error. If by the reliability coefficient we mean the split-half r and by the consistency coefficient, the test-retest, then function fluctuation is the excess of the error in the consistency coefficient over the reliability coefficient, since the latter expressed the variance due to error of measurement alone.

practical prediction from using factors are initially not as great as they later become, but even the pioneer has his rewards in economy and cross reference and is safe so long as he estimates his factors immediately in the situation in which he proposes to use them.

One particular reward in practical prediction not yet mentioned appears when criterion and tests are factorized together. Without factor analysis it is possible to add tests to a battery, each of good validity, without adding anything to the fraction of the variance of the criterion that is successfully predicted (except through increasing test reliability). The new tests may be involving only factors already unknowingly taken into the tests used in predicting the criterion. Factor analysis enables one to recognize how much of the valid variance is already being predicted and where one needs to look in wider varieties of tests for better measures of factors which have yet to be taken into account in predicting the remaining variance of the criterion. If this and one or two other general principles in these paragraphs are somewhat obscure to the reader at this point, he will perhaps excuse the writer their premature introduction in the interests of theoretical completeness of the present discussion and will return to them after the next few chapters have clarified whatever concepts are obscure to him.

The technique in applied psychology to which factor analysis eventually leads is one in which the practitioner basically needs only two files: a file of persons, with their standard scores on a small number of primary personality factors; and a file of performance situations, recording the situational indexes which have been found by research for the factors in various important real life performance situations, e.g., recovery in clinical therapy, success in various occupations, facility in various school subjects. In sociology and economics the factorial approach is too recent for us to illustrate even the general nature of the factors that would be used (except for the dimensions of national culture patterns) ; but in psychology, even though a comparatively small number of situational indexes are yet fixed, at least a substantial number of personality and ability factors are delineated and measurable for understanding the individual personality.

Questions and Exercises

1. What are the two principal uses to which a rotated factor matrix can be put?
2. Discuss in a preliminary way the sources of error which will normally have entered in the final, rotated factor matrix?
3. Write out the generalized specification equation for a test which cannot be accounted for wholly by common factors.
4. Discuss the problem of weighting test performances in finding an individual's endowment in a factor and indicate why the process is called estimating.
5. State the specification equations for variables 2, 4, and 6 in Table 10 of the previous chapter.
6. Write the specification equation for individual No. 39 on a group of tests among which 8 common factors were found.
7. Using the standard scores of persons A and B as used in this chapter, compare the performance of these two individuals on test 3 and test 6. In which test does A have a better performance rating? B?
8. Summarize under some four pro and two con items the advantages of factor analytic prediction relative to the practice of predicting criterion scores directly from particular tests not factorially pure.

CHAPTER 7

Unitariness in Relation to O-, P-, Q-, and R-Techniques

The most fundamental steps in factor analysis itself, such as the derivation of factors from correlations, the rotation of factors for scientific meaning, the estimation of the ensuing factors for individuals and the use of the specification equation for prediction, have now been described. Equally basic matters concerning the setting of factor analysis in different experimental designs and the forming of scientific conclusions and concepts still remain to be discussed. The more general of these issues concerning the scientific meaning of results from different settings can be dealt with in this chapter, but there will also remain the finer points to be postponed until the detailed mathematical issues which we have brushed aside in part one of this book have been systematically examined in the second part.

FACTOR UNITY IN THE LARGER CONTEXT

So far we have based our quest for scientific reality in a factor largely on its meeting the criteria of simple structure. The assumptions of this criterion have been stated in Chapter 5: that we should not expect any one psychological attribute to have major effects on more than a fraction of the total personality; and a parallel conclusion can be maintained with respect to independent social influences and the mass of sociological variables. Though the principle may be slightly differently stated by different experimenters, as will be seen in Chapter 14's discussion of the details of rotation, the common feature of the theorem upon which most agree is that the best hypothesis regarding factor structure is one which gives *parsimony of explanation* of the single given correlation matrix.

However, the present writer, in suggesting the *parallel proportional profiles* principle (**18**) shortly to be mentioned, has argued that the general scientific principle of parsimony is best applied in factor analysis *with respect to a whole population of factor analyses rather than a single matrix.* In other words, the question is not "Which factors give the simplest explanation of these correlations?" but "Which factors give simultaneously the simplest explanation of the correlations in this matrix and those obtained in many other experiments?"

This states in terms of correlation matrices what might be expressed verbally by saying that the functional unities in one factor analytic investigation should be the same as those found in another and that both should correspond to patterns of functional unity found also in simple experiment and natural observation in other contexts. This is exemplified, for instance, when the tests which are highly loaded in the general ability factor prove also to be those which distinguish themselves from educational measures by coming to a maturation plateau around 14 years of age and which show impairment by brain injury proportional to the volume of the lesion and regardless of its position.

A recent example from physical data of this demonstration of the pattern found by factorization in some quite different context occurs in the work of Cureton (**42**). He intercorrelated about thirty measures of heart action, circulation, and other physiological variables for a large, male population and found among four or five other factors one characterized by high loadings in pulse rate after exercise, diastolic pressure, expiratory force, and some muscular measures. Later he found that this set of variables—which he called the cardiovascular efficiency factor—separated themselves from others also by greater change under training conditions. They also separated themselves by showing more change in a low-pressure oxygen chamber.

TEST OF FACTOR REALITY

The reality of the functional unity which we call a factor can be thus tested both within and without correlation methods. Sometimes the noncorrelational demonstration may depend, as in the last illustration, on the variables in the factor showing each a significantly greater change than do other variables in response to some external influence, but also on many other relations of measurement, such as are implied

in differences of comparative growth, contiguity, common history, etc. These ancillary signs of the pattern may be known before or after the appearance of the pattern in factor analysis. Such checks lie outside factor analysis and need no further description or amplification here. But the checking of a pattern by the detection of similar factor patterns in different correlation studies is part of the factor analytic method itself and requires discussion of techniques.

When the comparison of factor matrices from different correlation matrices is mentioned, we necessarily refer to correlations among the *same battery* of tests, or at least batteries overlapping in test content, when applied to different samples of people. Needless, perhaps, to add, in any test of a hypothesis of similarity by this means, the second analysis must be rotated independently of and without guidance from the first. For, as will be seen, it is sometimes possible to produce a tolerable imitation of one factor by another—at the cost of some confusion of other factors—by deliberate rotation. To insure unquestionable independence of rotation, the variables must be shuffled and represented by numbers not known to the person rotating for simple structure, in a process which we shall hereafter call blind rotation. Under such conditions it is assumed that the rediscovery of the same factors despite (a) partially different test batteries or (b) populations of different age, education, or dispersion, and (c) independent factorizations and rotations, is a proof that they have an existence as something more than mere mathematical equivalents—that they are in fact functional unities in nature.

O-, P-, Q-, AND R-TECHNIQUES

Such a test of factor reality is reasonably satisfactory, but there is a still wider sense in which constancy of pattern may be demanded and tested. This requires a comparison of the factor patterns resulting from methods which have been called O-, P-, Q-, and R-techniques in factor analysis. The great majority—perhaps 95% of all factor studies to date—have used R-technique, the correlation of variables (two at a time) using a series of persons as entries (points) in the correlation. Though the transposed factor technique, called Q-technique and devised by Burt (**9**) and Stephenson (**117**), will appear very simple and obvious when pointed out, its possibility was actually not noticed until after a decade of factor analytic experience. It consists in nothing more than correlating persons instead of correlating varia-

bles, i.e., in looking at the usual correlation table from the side when seeking columns for correlation instead of from the bottom as we usually do. This is illustrated in Table 11 where the same scores are shown used for two different kinds of correlations.

TABLE 11. Correlation Series in R- and Q-Techniques

	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	
Adams	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	←These two
Brown	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	←rows correlate
Carruthers	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	to give an r in
Cox	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	Q-technique.
Doe	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	
Jones	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	
Robinson	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	
	↑	↑							

These two columns correlate to give an r in R-technique.

From working out all the correlations among the rows instead of among the columns, we shall obtain a correlation matrix *among people*. Usually, in order to get r's of ample statistical significance we take a large population of people and a reasonably small group of tests. For Q-technique the converse holds; we do not particularly need to insure many people, *but we must have many tests* if the r between any two people is to be reliable.

What does this r between two people mean? Clearly it indicates the extent to which the two people resemble each other with regard to the series of tests in question. This point is most easily illustrated by taking a series of ranked items, say pictures, instead of tests. If Smith's ranked order of preference for the pictures agrees well with Jones's, we infer that, at least in artistic tastes, their personalities are alike; whereas if the orders are opposed, the correlation of Smith with Jones will be negative.

A correlation matrix of people may, and often does, show clusters just like those in matrices of variables. Each constitutes a group of people who are all alike but who have no particular similarity to those in some other cluster. Q-technique is thus an ideal method for finding

types if such types actually exist with respect to the variables concerned. The individual who shows the highest mean intercorrelation with all others in the cluster is the most perfect representative of the type.[1]

By contrast with R- and Q-techniques above, which involve a population of persons, the P- and O-techniques apply factorization to a *single individual.* These methods, first suggested by the present writer in 1946 (**22**), were intended primarily for clinicians wishing to replace free association by a more powerful and positive method of analysis, and for research work upon the structure of the self-sentiment (**30**). First demonstrations of these usages in a variety of fields, from clinical psychology (**23, 30**) to anthropology and history (**27, 32**), have been given by the writer and his students: but these are trivial beginnings of what has potentially a much vaster application. The manner in which correlatable series are set up for the single individual is explained in more detail in the sections on P- and O-techniques below.

DEFINITION INCLUDES DESIGN, SCALING AND DATA

At the time of this writing, clinical psychologists are showing a belated zeal for what factor analytic techniques can do to bring to clinical psychology the scientific method which was naturally sought in vain by classical, controlled experiment. Unfortunately, this precipitate enthusiasm, especially when distorted by the accident of some crusading imperialism among one or two followers of Q-technique, threatens to replace disuse by misuse. Consequently, we must examine, with a little more space than might otherwise be given to a mere examination of pitfalls, some of the misleading statements and assumptions that have been made in relation to Q-technique. At least the necessity for this consideration may have the fortunate result that the proper relations of R-, P-, Q-, and O-techniques will be brought to a more explicit level.

[1] The concept of type, as shown elsewhere (**22**), can be used in three senses—as continuous type, discontinuous (bimodal) type and species type. Although a continuous type may be designated by a cluster, as here, it may also be defined by a cluster or factor in R-technique data. For example, we speak of a mental defective as a type designated by all the manifestations which form the loading configuration of *a single factor*—low intelligence. Whether the cluster or factor found in Q-technique does or does not correspond to a factor in R-technique will depend upon whether the range of the tests used in the Q-technique study extends over many factors or one.

In particular, it has been claimed by Stephenson that Q-technique permits one to obtain scientifically valuable results with much smaller populations and far less calculation than other techniques; that it yields knowledge of individual personality structure not inferrible from other methods; and that it is less subject to certain distortions. He has particularly urged a method in which each subject writes down a set of traits or questionnaire-like statements about himself in order of their *significance* for his own personality (*Q-sort*). This order is then made the basis of a rank correlation with another personality similarly self-rated.

The waste of research effort which has supervened arises basically from confusion as to the differences of R-, P-, Q-, etc., techniques. These techniques are basically variations of experimental-statistical *designs*, generally with variations in *scaling* procedures and variations in *universe of data observation* added. These three aspects of a research are essentially independent, though occasionally certain combinations are not possible. In general, however, R-, P-, Q-, and O-techniques keep their form regardless of whether the scores have first been normalized or otherwise scaled, and regardless of whether they operate upon data gathered by observing behavior *in situ*, or by self-rating (introspection) or by objective tests. Of course the results of the different combinations may have very different validity and field of true reference, and these we must examine.

The third matter, the level or *universe* of data observation, can be dismissed first, for it is comparatively simple and has also been treated sufficiently elsewhere (**22, 30**). Before any test or observation is "validated" or brought into relation to some other variables, it is usual to test its reliability by a reliability coefficient. Now reliability coefficients are of two kinds, only one of which has universal value.[2] When we take as datum a piece of behavior in real life or an objective test, it is observable and checkable by as many witnesses as we wish. A reliability coefficient between two such objective, publicly witnessable, behavioral measures we will write r_B. On the other hand, when a person answers a question about himself, by introspecting (e.g., Am I a cheerful person?), no one else can witness the truth of it. A reliability coefficient between two such observations restricted to one witness, who stands in a biased relation to the data, whether it concerns feeling

[2] If we admit that objectivity, valuableness and verifiability of data are of primary importance.

or behavior, because he is uniquely involved, we will call r_s the *S* indicating subjective, introspective *data*. Incidentally, questionnaire data is S-data only when the answers are accepted with their normal conventional symbol meaning. Some questionnaires, e.g., the 16 P.F. (**31**), avoid this and are behavioral. But generally a questionnaire gives S-data. A third kind of reliability coefficient is sometimes attempted, namely, that between the objective and the questionnaire data, where one observer has inside bias and the other has not. But this is incorrect, for, by definition, a reliability coefficient is between *similar*, "twin" measurements. This last is really a validity coefficient —providing each of the observations, the behavioral and the mental interior, have first themselves been assessed for reliability against a true "twin" measurement.

Although the universes of observation are thus in essence two only, namely, behavior and consciousness (introspection), corresponding to two varieties of reliability, there is so important a difference between two of the kinds of behavioral data that the present writer, in various factorial studies, has suggested keeping them separate, thus making three kinds of observation in all. The split is that between *life record* behavior, i.e., behavior actually in the everyday life situation, e.g., number of automobile accidents per decade, and *objective test* data, i.e., measured behavior in a controlled, portable situation, e.g., cancellation speed, reaction to threats on a psychogalvanic apparatus. Consequently, personality factors in at least a dozen published researches have been classified as *BR* (behavior rating *in situ* or life record, *LR*) factors, *Q*, or *S* (questionnaire or introspective, self-rating) factors and *OT* (objective, behavioral test) factors. It would greatly assist clarity in discussion of factor systems to adopt a nomenclature in which these three sources of data—or the two basic sources, behavior and introspection—are indicated by a subscript to the main design symbol. Thus we should speak of R_B, R_S, and R_T to indicate R-technique on each of these bases, or, if we wish to have a dual division only, in regard to, say, the factorization of persons, we could speak of Q_B-technique and Q_S-technique.

The clinical use of Q-technique has been entirely confined to Q_S-technique, and in regard to this, Sir Cyril Burt and others (**11**) have rightly objected that the instruction to "Rank according to the significance to your personality" is differently interpreted by different subjects. It is not that they are arranging traits in different orders,

according to their self-views, along the same continuum: The very continuum is different, because "significant" is a loose verbal symbol differently interpreted by different people. Whatever instructions one gives subjects it is always possible to get "results"; but these results are much less scientifically useful than they might be if the stimulus situation were more exactly defined and free of ambiguity. All tests using words with introspective reference, either in the instruction or the response, are liable to a source of error which may be styled "introduction of unwanted variance through uncertainty of symbol reference." Thus in Q-data, when a man replies "very much so" to "Do you consider yourself hard up?" he may mean in *BR*-data terms that he has anything from five to fifty-thousand dollars and, in introspective terms, that he is jokingly aware of financial limitations or positively obsessed by financial anxiety. The universe of Q-data, as argued elsewhere, is scientifically poor, but it is doubly so if incorrect instructions are given. As will be seen in a moment the only instruction that is statistically defensible for Q_S-technique is to ask the subject to rank or rate the traits according to *the extent to which he deviates from the mean*[3] *therein*, i.e., ask how outstanding, in terms of eccentricity, the traits are in his personality—and this is not giving him an easy task!

Let us now examine the question of scaling. Various aspects of scaling will be met at appropriate points in this book, and we shall confine ourselves here to those which alone are relevant to controversies over R-, Q- and P-techniques, namely the questions over the desirability of putting raw scores in standard scores. Stephenson (**117**) has written of four scaling systems, calling them "factor systems" and "the foundations of psychometry." The problem of scaling systems will be readily grasped if we start with a simple physical rather that psychological example. Let us suppose we have measured eighty men upon one hundred variables, each concerning some aspect of physique, e.g., height, weight, length of leg, breadth of hand, etc. If we take the matrix of scores and factorize by R-technique, our first factor will probably be one of *general size*, for the man with greatest stature will *tend* to have greatest leg length, greatest hand breadth, etc., so that from the resulting substantial positive correlations of the variables a factor not unlike general intelligence among abilities will arise.

[3] His own for all traits or the populations for that trait, according to which scaling we intend later to apply.

If now the same score matrix is handled by Q-technique it will first be turned on its side, so that the columns to be correlated are people and the rows are tests, as shown in correlating the a's with the b's in Table 11. But these columns will have a strange look, for *all* persons will have stature as their numerically largest variable (unless we have put weight in ounces!) and perhaps the length of the little finger as the smallest. All people will correlate highly together in what we may call a *common species* factor, defining *humanity* and measuring various individual resemblance to the common man. This factor is irrelevant for most differentiation within the species *homo sapiens*, and only after its extraction shall we come to the factors of individual differences in which we are interested.

One way of getting rid of this factor is to standardize each test row before we start the correlation of the person columns, i.e., to assume the tests have the same means and the same sigmas. Then the highest score for Brown will be that measurement, perhaps the length of nose, in which he departs most from the average of all noses. However, different *individuals* will still have different averages, e.g., though Brown's nose is relatively the biggest part of him he may still be below the mean on all his scores. And now, since the correlation coefficient (but not r_p (**26**), the pattern similarity coefficient) gives a correlation of unity for two series of scores which are perfectly parallel but not at the same level, a small man, like Brown will correlate highly with a man of similar physical profile, even though he is much larger. In short, *Q-technique will give a* species factor *but fail to register the* general size *factor which would first appear with R-technique on the same data,* and if we first *put the tests in standard scores it will also omit the common species factor.*

The argument for *four foundations*, actually four alternatives in scaling, states, according to Stephenson (**117**), that the following alternatives are possible.

1. The data is rescaled by standardizing it for the population of persons. This means that the tests are brought to the same mean and sigma, on the assumption that in the general population, if not the sample, there is no sense in saying, for example, that people the world over are taller than they are heavy. In other words the particular units in which various measures are taken, e.g., whether centimeters or inches, are unimportant, and there is no meaning to differences of means of tests. This is the basic position taken in the present work. In

R-technique the rescaling does not, in any case, affect the result; but as we have seen in the example above, it helps to eliminate in Q-technique an irrelevant effect. The procedure if applied to the scores in Table 11 would mean adding and finding the sigma of all the values with sub-1, and the same for the sub-2's and so on.

2. The same procedure may be carried out for persons instead of tests, i.e., for all the *a*'s, all the *b*'s, etc. in the raw scores in Table 11. Unless the tests have first been taken out of their initial, accidental raw score form, as in (1) above, this procedure seems to the present writer useless for any factor analytic purpose. If applied to R-technique it would remove the general size factor, but the fact that people differ in general size is something we want to know.

3. Raw scores may be rescaled as in (1) and *then* rescaled as in (2). Like (1) this is recommendable for Q-technique, but the second steps of bringing all persons to the same mean and sigma loses some of the information available—at least if we use the pattern similarity coefficient (**26**) r_p instead of R.

4. The last logical possibility is to reverse (3), carrying out (2) before (1).

Stephenson argues for (1) and (4) as being suitable for R-technique and (2) and (3) for Q-technique, but, as seen above, *any* scaling procedure can be used with *any* design and *any* data. The usual effect of standardizing is to lose some information in the data—in fact to lose a factor from the row correlations when the columns are standardized, and vice versa—so that it should be avoided except where the information is irrelevant or misleading.

The two last of the above four scaling systems are in any case superfluous, because correlation automatically *standardizes* along the series being correlated, i.e., it neglects differences of mean and sigma in the columns when we are correlating columns, as seen above. Consequently there are really only two alternatives to be considered in scaling, namely: (1) to leave the data in raw score form; or (2) to standardize it *across* the direction of correlation, i.e., for rows when correlating columns and vice versa. The second has utility in eliminating the *species factor* in Q-technique, but is pointless for most R- or P-technique.

However, for clarity of future discussion, let us designate the combination of design, scaling, and data by a capital and two subscripts, in that order. Thus R-technique carried out with scores scaled by method

(1) upon behavior rating data would be R_{1B}, while Q-technique on subjective, questionnaire data using scaling method (2) would be Q_{2S}.

THE RELATIONSHIP OF TRANSPOSED TECHNIQUES

When used on the same data and with the equivalent scaling, Q-technique has been called the *inverted (or obverse) factor analysis* with respect (**117**) to the R-technique. As Burt has pointed out (**9, 11**), it would be more appropriate to call it *transposed*, for the score matrix is the *transpose* of that used in the other method, and we shall adhere to this improved nomenclature. Just as Q-technique is transposed R-technique, so O- is the transpose of P-technique and S- of T-technique.

Now Sir Cyril Burt has argued in relation to R- and Q-techniques —and the same would hold by extension to the other two pairs—that except for certain special modifications they yield the same factors. If correlation of ability tests yields, after the extraction of general intelligence, a factor of mechanical aptitude, then correlation of persons will yield a corresponding factor of "mechanically apt persons," the only difference being that in R-technique we point to a highly loaded *test* as exemplifying the factor, whereas in Q-technique we point to a highly loaded *person*. It seems reasonable that if we start with the same matrix, merely turning it on its side, we should finish with factors which (except for the first factor mentioned above), would be the same after rotation, i.e., which are at least mutually transformable. Burt's criticism of Stephenson's claim that the method which Burt and Stephenson invented is statistically independent of R-technique should be carefully read by those now assuming their independence (**11**). Moreover, the statistical assumptions of independence have been recently examined more closely by Madow, who point out that R- and Q-techniques yield, from the same score matrix, (1) the same number of factors, and (2) the very same factors, when the scores are scaled to be "doubly-centered," i.e., when rows and columns are standardized to the same mean and sigma. Even if this exact equalization is not achieved the usual departures from it (through standardizing row *after* columns) are too slight to affect the essential similarity of the two factor structures derived.

Statistically, we may generalize that each technique and its transposed form stand in an entirely symmetrical relation. Thus, in regard

to R- and Q-techniques we correlate in one the columns, in the other the rows; in one we neglect the differences of means between tests and in the other the differences of means between persons, and so on. But this reciprocity breaks down in the case of R- and Q-techniques at five points—some statistical, some experimental—which no argument for symmetry must persuade us to overlook, and in respect to R- and Q-techniques on social science data at least four of these discordances are to the disadvantage of Q-technique.

First, Q-technique loses the "general size" or first common factor, because, as seen above, persons of the same pattern but different size are considered identical by the correlation coefficient. R-technique, reciprocally, fails to find the first factor found by Q-technique. But this factor of *common species* is unimportant, since the experimenter knows he is dealing with human beings and, except for special purposes, requires no statement of the species character. Occasionally, as stated, the species factor is of interest, e.g., we may need to be reminded that men's legs are longer than their fingers, but in terms of psychological tests the fact that people gain more raw score points on test A, scored according to such and such a scheme, than on test B, etc., is a triviality of test construction.

Secondly, the symmetry breaks down with respect to rotation for simple structure. For, whereas it is reasonable to assume that some tests will completely lack loading in certain factors, permitting a hyperplane, it is difficult to argue that some human beings will *completely* lack some factor present in others. In the handful of Q-technique studies yet performed simple structure has generally not been sought, but to the writer's knowledge it has only been found in one or two wherein a *nonhomogeneous* population was used, e.g., artists and non-artists, so that a certain factor (artistic training) could indeed be totally absent from many persons. Thus simple structure seems applicable only in heterogeneous populations of several species and elsewhere the experimenter is lost in an arbitrary rotation. However, the usual normal distribution of factor endowments in persons would cause more to stand at zero standard score than elsewhere and thus give guidance through a pseudo-simple structure.

Thirdly, a score matrix is usually oblong, because experimenters have to take many persons, to make the correlations reliable, and whereas this may lead to two hundred or more rows, few are prepared to intercorrelate two hundred or more columns! To use the transposed

technique requires, therefore, either that we enlarge the matrix to a terrifying size or that we reverse the frequencies and have many tests and few people. This may at first seem a fair and symmetrical exchange, but if we are interested in generalized scientific findings, applicable to the total human population rather than to the sample only, *it is important that this sample be large enough to permit extension of the findings to the population* with little sampling error. This means some hundreds of cases rather than the dozen or so with which clinicians are claiming the right to employ Q-technique.

What of the converse shortage—the fewness of tests in R-technique? This could be and often is a restriction of the factor picture, and for this reason the writer argues in later chapters on research design for (1) larger matrices with more randomly or eclectically selected performances and (2) a conception of the total population of tests—the *personality sphere* (**22**)—equivalent to the "total population" of persons. However, some researchers will not agree that there is any exact equivalence between the notion of a population of persons and one of test performances and they consider there is no onus upon them to sample tests. This much can be said in defense of the position: That tests are preservable and persons are not, wherefore it always suffices to say "These are the tests in which I found the factors named, and they can be found in them again by employing the human population." At the same time it can be pointed out that the mere use of a large number of tests in Q-technique does not guarantee a sampling of the test population, as a large number of people normally guarantees sampling of the parent person population. Tests are more idiosyncratic.

In fact in the few Q-sort studies reported, in which Smith is said to resemble Jones highly because they correlate highly in the order in which they have ranked features of personality for significance, there is not the slightest guarantee that they really resemble each other highly. For *all* the items may have to do with some small corner of personality, covering only one or two primary personality factors—in one actual study they all dealt with choices in modern art—and the individuals may differ greatly in all the other personality factors. Indeed, to carry out Q-technique or P-technique effectively it is necessary to use tests sampled from the principal personality factors already found by R-technique employing the *personality sphere* concept.

Fourthly, there is a lack of equivalence in recording and interpreting of factors. By R-technique we finish with some highly loaded tests, which can be preserved in files for record and the operations in which can be observed in any member of the population, in order to understand what the factor is. By Q-technique, on the other hand, we finish with a highly loaded person, who is changeable, perishable, and not susceptible to filing. Of course we may file his profile of scores on all the tests at once, but this is a more bulky proposition and the task of interpretation from a profile is far more difficult than interpretation from a single test process. An alternative suggested is to interpret from the person himself, but this is even more chancy than the profile. For example, if we take a person highly loaded on one factor deemed to be intelligence and with zero (average) loading on all other factors, it is hard to infer from his behavior what is the nature of intelligence. He is also sociable, competitive, and a thousand other things in his living acts, so that observation of his general behavior will not clearly show the nature of intelligence. On the other hand, examination of his profile will show him below average in tests a, b, and c and above average in l, m, and n, from which, somehow, we have to infer the factor. By R technique we should have, say, a couple of tests, analogies, and classification, which are highly loaded in the factor, and inspection of their common character, namely, *relation eduction*, would show the nature of the intelligence factor.

Fifthly and lastly, the techniques are reciprocal in relation to the practical problems or operations which they more immediately aid. Q-technique is most useful if one wishes immediately to see how many types there are in a population and to divide it up into types. This usually has merely descriptive value. On the other hand, R technique leads more immediately to the use of the specification equation, whereby the performances of the individual in a great variety of specific situations are predicted from his factor endowment.

The decision as to use of R- or Q technique must sometimes be left to circumstances. Whenever one can get many subjects, R technique is preferable. If it is absolutely impossible to get more than ten to twenty subjects, one is compelled to use Q-technique and test each on a much longer series of tests. It is obviously a mistake to suppose that Q-technique by using fewer subjects saves time. The subject hours of observation remain about the same since in taking fewer subjects

one requires more hours on each. Since much testing is group testing, R-technique, with many subjects, means less of the experimenter's time.

Moreover, one must also not fall into the error of supposing that in some magical way the findings on very small groups can be generalized as well as those on large groups. Sometimes Q-technique enthusiasts avoid this criticism by saying that they are interested only in the relationships within the small group. If this is sincerely and meticulously followed it is correct, but almost invariably, the investigator is not really prepared to restrict his generalizations to the little universe, e.g., a particular family, which he seeks to understand. He inadvertently generalizes from the types and factors he finds in the family or a couple of clinical cases to the structure of personality in general. But even if no false attempt is made to contribute to general principles one must still point out that even the understanding of the restricted universe is not possible without reference to the general universe. The poet has said: "What know they of England, who only England know?" and one can similarly point out that the relations found by Q-technique or P-technique in a few cases or a single case can only be fully understood in the light of similar or dissimilar structures in the population generally (or by reference to a whole series of Q- or P-technique researches). And in general, R-technique is the binding frame of reference through which all the other techniques are brought into due reference to the general population of people and generalized psychological processes. Consequently, most people who come to have experience with Q-technique will probably decide (once they recover from the specious impression that it will yield something for nothing or, at least, for greatly abbreviated labor) that the overall advantage definitely lies with R-technique; and it seems likely that, except where special circumstance, e.g., need for rough exploration of a wide area, exist, R-technique will continue to be the generally preferred tool.

P-TECHNIQUE

The third method of using factor analysis—P-technique—differs more from the above two methods than they do from each other. They are reciprocal designs in relation to the same experiment, but P-technique is a totally new experimental and statistical situation. It begins by measuring a set of variables on *one person* and repeats

these measures on a sufficient number of occasions to provide correlatable series. It will be clear from Diagram 13 that those variables —e.g., A, D, and E on the one hand, or C and F on the other— which tend to change with time and circumstance *in the same way* will emerge as a correlation cluster in the matrix. By factor analysis we can then obtain the independent influences or dimensions of fluctuation and change. Incidentally it is unnecessary to eliminate an overall trend from the data, as some social scientists have proposed to do in time series, e.g., in economic material, in order to reveal

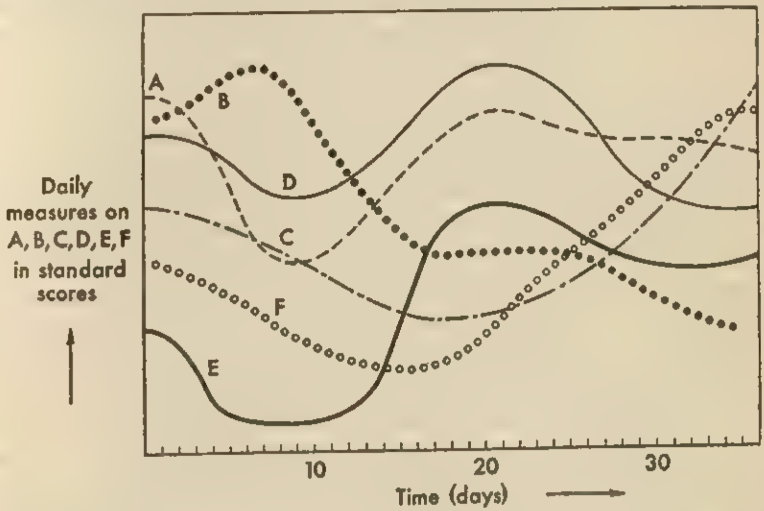


DIAGRAM 13. Trends Used in Correlation for P-Technique. (From *Description and Measurement of Personality* by R. B. Cattell. Copyright 1946 by World Book Company. Reproduced by permission.)

the other connections more clearly. Factor analysis will itself partial out and set aside a *trend factor* if it exists; for factor analysis is essentially the working out of partial correlations. The loadings are correlations between the variables and the factor when other factors are held constant.

P-technique can be carried out with or without deliberate experimental attempts to produce fluctuations in the measurements. Thus its use in psychology may or may not be accompanied by controlled change of the stimulus situations such as would cause certain source trait patterns, e.g., that of hunger, to fluctuate strongly as a single

whole. Some psychologists have been prone thoughtlessly to assume that the day-to-day variations in measurements on an individual are largely experimental error of measurement. It has recently been demonstrated (23, 141) that at least six of the primary personality factors fluctuate sufficiently under the natural impact of daily events to yield significant *r*'s among their parts, and it could well be that with extension of the number of occasions of testing from 100 to 200 or 300, the increased significance given to the lower *r*'s would permit all factors known from R-technique to appear.[4] In the event that actual manipulation of stimulus situations or internal physiological conditions is employed in P-technique, these influences should be correlated in, as variables in the matrix, in order that the instrumental relationship to the traits may be demonstrated.

The use of P-technique does not require that the occasions be equally spaced or that periods be as short as diurnal or hourly intervals, but the interpretation of time-related factors, e.g., of fatigue or learning in psychology, of climatic or sunspot cycles in social data, is more reliably undertaken if they are.

In psychology, P-technique has at the moment occasioned greatest interest in connection with (a) the possibility of exploring unique traits (22) in the individual where previously the only uniqueness assignable to the individual was a uniqueness of combination of common (R-technique) traits, (b) the readiness with which it permits exploration of psychosomatic relations by bringing physiological and psychological observations on the same organism into a single matrix, and (c) its possible supersession of free association and similar methods in clinical psychology or wherever a more positive methodology is required for exploring the dynamic connections in the single organism.

P-TECHNIQUE WITH STAGGERED TIME SERIES

But it may prove to have an even greater value to sociology, economics, and history, for the factorization of time series offers an objective analysis of historical causation and a definition of independent movements not yet attainable in any other way.

A special development of P-technique is that in which all the

[4] The question of whether ample function fluctuation occurs must be decided from a comparison of the reliability and consistency coefficients.

measurements for one variable are staggered in time (or in occasion sequence) relative to the measurements of another variable on the assumption that there is a time lag in their mutual influence. If, for example, we obtain a factor connecting poorness of memory and amount of alcohol consumed, and suspect that the latter is the cause of the former, we may raise the correlation in the time series by correlating Sunday's memory score with Saturday's amount of alcohol consumed. By staggering the series by a one-day lag in this fashion we may actually increase the correlation initially noticed and raise the loadings in the factor expressing the interconnection.

This appears again as the problem of lead and lag in business cycle data, but is common to all the human and social sciences where delays of communication of influence as well as various feedback mechanisms (**140**) blur the picture of factor loading as obtained by uncorrected P-technique. For in those situations, e.g., R-technique, where time is ample for causes to produce their effects, it is possible to argue that the variable most highly loaded in a factor, and therefore ultimately the factor itself, is the *cause* of the manifestations presented by the lesser loadings (see Chap. 16). At present no method of calculation is known for determining the lag which will make the r between two series a maximum—it has to be discovered by toilsome trial and error.[5]

P-technique does not yield factors having the mathematically necessary connection with those of R-technique which Q-technique factors have when established on the same group. It is, in a sense, an R-technique study (in that it correlates *variables*) on a different—a radically different—sample. But essentially it is to be regarded as a truly independent method, using a different population—one person—and different units, namely *ipsative*[6] units instead of normative units (**17, 22**). It yields the unique patterning of each source trait as it exists in the given individual, but the evidence so far is that these unique factors turn out to be slightly distorted copies of the patterns of common factors obtained by R-technique. In any case, we should expect this degree of resemblance if the factors correspond to real influences.

[5] There are, however, mechanical devices, such as cutting out profiles for the two series and sliding them along till a minimum interference with transmitted light occurs.

[6] Units in which the raw scores have been expressed in standard scores with respect to the σ of the population of occasions, i.e., of fluctuant measurements within the individual, instead of with respect to a population of persons, as in normative scores (percentile and standard scores).

O-TECHNIQUE

As in P-technique we deal now with *intra-individual* factorization of a single person or organism. But O- is the transpose of P-technique as Q- is of R-technique. That is to say, after building a score matrix by measuring the same person on, say, two dozen different tests on each of one hundred days, we correlate the *days* instead of the *tests*. In its generalized form this means the correlation of *occasions*, whence the symbol O is as appropriate as is P for the single *person* study.

Both O- and P-techniques are peculiarly suitable for clinical psychology and the study of the self and its integration. They also have special application in the social sciences, notably social psychology, economics, and history. In clinical psychology P-technique, as indicated, is the most positive method available for establishing connections in individual dynamic expressions and symptoms, whereas O-technique, as pointed out by the writer in 1946 (**22**) offers an ideal method for investigating multiple personality. For in the latter phenomena the essential situation is that though there are many occasions of observation only two, three, or more "selves" are in action. By the correlation of all occasions in a sample "population of occasions" we may expect to obtain a correlation matrix in which certain groups of occasions cluster together and perhaps yield three or four factors. Each factor can be recognized by a certain pattern among the two dozen test measures and that pattern constitutes a relatively stable personality or self among the possible patterns from moment to moment. As in Q-technique it is a methodological improvement to have the variables in the pattern strategically sampled, either with regard to some special hypothesis about the self or to represent the important factors already found by R-technique.

O-technique studies are also in process on longitudinal studies of nations or societies, whereby the pattern of culture on one occasion can be compared with that on another, thus establishing the degree of reality to be attached to concepts of era, epoch, historical phase, business cycle phase, etc. In this as in P-technique it is naturally assumed that the different occasions will differ in some significant respect. Indeed, each occasion may amount to a different stimulus situation and may even be made so by experimental control. Therefore each occasion should be labelled and described as fully as possible by its features. For the different occasions in O-P-techniques are the

equivalent of different tests in R-Q-techniques and are the basis for defining the factors. Thus in multiple personality studies it may become evident that one personality factor pattern loads, i.e., tends most to appear upon, those occasions when the individual is visited by his parents, another in occasions of intoxication, and so on.

Since the relations of O- and P-techniques will be precisely the same, *mutatis mutandis,* as those of Q- and R-techniques, nothing essential needs to be added to the above discussion of R-Q transposition and reciprocity. In this case it is O-technique which turns out to be somewhat the less manageable and widely applicable of the two. For it requires that a large number of tests be given: it defines the factor (via the occasion pattern) rather less definitely, and it substitutes for the omitted first factor among tests a first factor among occasions—"the person as he is on the typical occasion"—which is less important, because, in a sense, it is already known.

The use of comparisons from different factorizations in an attempt to prove the reality of factors is developed in more comprehensive and systematic fashion in the next chapter.

Questions and Exercises

1. Describe three basically different sources of confirmation for the reality of a factor pattern found by a particular factor analysis.
2. Define the general nature of the variables and the population in R-, Q-, O-, and P-techniques.
3. In what three senses can the notion of type be used and what relations can you see between any of these type notions and the products of Q- and R-techniques?
4. Describe and illustrate by drawing a score matrix the difference between O- and Ps-techniques. Comment on the meaning of measurement in these applications.
5. What are the chief characteristics of a research problem that you would take into consideration in deciding between R- and Q-technique designs?
6. What difficulties in Q-technique (a) practically invalidate it for common purposes of factor analysis and (b) are only of a minor distorting or complicating effect?
7. Give a relatively detailed account of the conditions and options in the use of P-technique and, in discussing the relation of P-technique to R-technique factors, comment on the meaning of unique traits.
8. Describe some fields in which P-technique is particularly useful. Discuss the nature of ipsative units of measurement and explore by examples the meaning of factors obtained from matrices using staggered correlations.

C H A P T E R 8

The Covariation Chart and the Possibilities of Obliqueness, Order, and Efficacy

Although the O-, P-, Q-, and R-techniques just described are the principal methods of factor analysis that have been developed to analyze the covariation observable in the chief situations encountered in practice, they do not exhaust the theoretically possible designs according to which it is possible to gather data, arrange it for correlation, and obtain factors.

THE COVARIATION CHART

Without dallying over any unsystematic exploration let us at once succinctly summarize the theoretically possible schemes by Diagram 14, which has been called the *covariation chart* and which crystallizes the results of many explorations. It starts with three parameters which are essential and sufficient to place or define any observation: (1) the time or occasion on which it is made, (2) the terms in which it is made, i.e., what variable is involved, and (3) the place, reference point, or organism of which the variable is an attribute.

This may be grasped readily if we take a particular field, namely personality study, where the time or occasion marks the moment at which the measurement was made; the terms of measurement become the scores of a particular test and the reference point is a person. If we are correct in supposing that these are the only basic conditions for making a measurement (a ruler, an object, and a moment) all the possible uses of correlation and factor analysis are contained within the combination of correlatable series obtainable from this model.

A correlation requires two series of measurements in which a value in one is always paired with a value in the other, through some

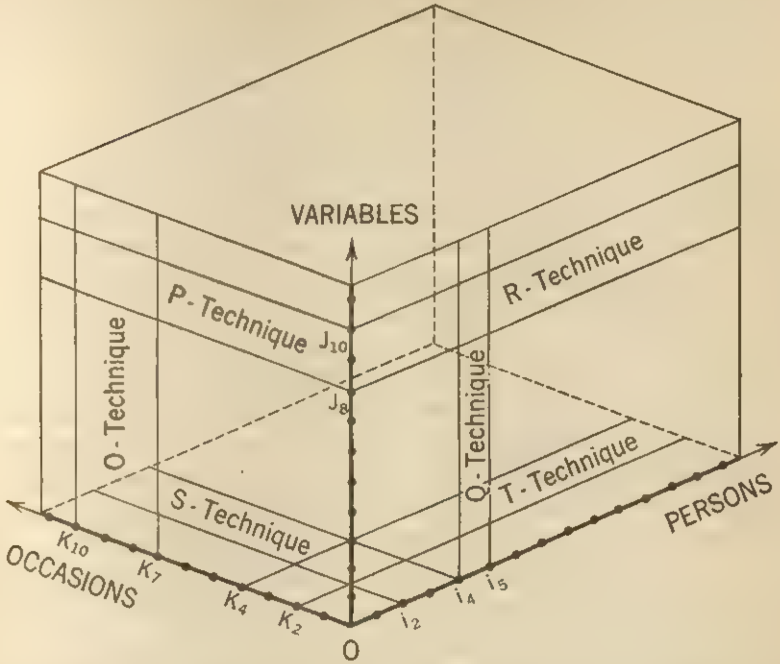


DIAGRAM 14. The Covariation Chart. (From *Description and Measurement of Personality* by R. B. Cattell. Copyright 1946 by World Book Company. Reproduced by permission.)

meaningful, systematic association binding each pair—in the same way for all pairs. The six strips cut in the covariation chart show six such correlatable series, corresponding to R-technique (two points in the series of tests continued in line through a whole series of persons); Q-technique (two points in the persons series tested on one occasion upon a whole strip of tests); P-technique (two tests correlated along a series of occasions); O-technique (two occasions correlated for one person with respect to a whole series of test performances) and two new designs not previously discussed which have been called S- and T-technique, the latter being the transpose of the former. These two logically possible but as yet unused designs correlate respectively two persons on one test on a series of occasions, and two occasions, on one test on a series of persons. The former has particular promise in *social* psychology—for measuring the similarity of response of two people, e.g., husband and wife, leader and

follower, twins, etc., on a whole series of social occasions—whence the symbol S is appropriate. Its factors will indicate sets or classes of people who react similarly on a number of occasions, thus its greatest value is for empirically determining groups and roles.

T-technique is nothing more than the factorization of reliability coefficients for the same test on the same people on many different occasions. The symbol T can be readily remembered if one thinks of the design as a collection of *test-retest* coefficients. The factors will show themselves as groupings of occasions which have some essential similarity in regard to their effect on the test performance. The design might thus be used for determining what and how many distinct elements (or "atmospheres") in the stimulus situation affect people's responses. It might also be used to detect similar atmospheres in social groups on different occasions. In this it resembles O-technique, but the latter has to deal with society as a single organism and must therefore take syntality (**25**) measures as its variables, whereas T-technique can work on measures taken on each member of the whole population of persons. It has immediate value in structuring the population to get greater meaning from opinion polling surveys.

To complete the systematic view of factor analytic techniques, it should be pointed out that there are essentially *three* experimental designs, leading to *six* techniques and *twenty-four* derived procedures. The three basic experimental designs are represented by the three pairs of internally transposable techniques, namely, R- and Q-techniques, O- and P-techniques, and S- and T-techniques. Each of these is represented in the covariation chart by one face of the parallelopiped (or any faces cut parallel thereto). Each holds one of the three attributes of a measurement event constant and varies the other two. *Thus R-Q holds to one occasion; O-P holds to one person and S-T to one test.* Incidentally, it is possible to perceive relatedness among the six when paired in other ways. Thus P- and S- share the use of a series of occasions; O- and Q- of a series of tests; R- and T- of the series of persons. Thirdly, among the possibilities of relatedness, P- and R- share the correlation of tests; O- and T- the correlation of occasions, and Q- and S- the correlation of persons. This last grouping, according to that which is correlated, is as important from the standpoint of integration of research as is the primary grouping according to transposability. Thus R- and P-techniques go together

in joint harness as the principal means of analyzing personality structure.

The derived forms arise from the alternatives of scaling and data universe discussed above. The raw scores can be used directly, in which case we may write R_1, Q_1, etc., or they may be standardized along the rows of the matrix when we correlate columns, and vice versa. In the latter case we write R_2, Q_2, P_2, etc., the subscript 2 indicating the second possibility of scaling. The six subscript 2 derivates will each lack one factor relative to the six subscript 1 factorizations.

The twelve derivatives thus reached are multiplied to the final twenty-four possibilities because each admits of two universes of data—behavioral, public data or introspective, questionnaire, consciousness data. The shorthand which permits accurate, unconfused reference to any of these twenty-four systems is that already advocated above, in which the first subscript, 2, or 1, indicates the scaling or lack of scaling, and the second, B or Q, indicates behavioral observations (behavior rating, BR, or objective test, OT, data) or questionnaire, self-rating, Thus R_{1B}-technique is the typical and widespread procedure in test factorization; Q_{2B}-technique is Burt's advocated use of Q-technique; Q_{1S}-technique is Stephenson's use, and P_{1B}-technique describes the method of the four studies published to date in the clinical analysis of the single individual.

These twenty-four categories constitute the minimum number to which the available *independent* systems can be reduced. For sub-B (behavioral) systems are in different universes from corresponding sub-S (introspective) systems; unscaled systems, sub-1, have a factor not found in scaled systems, sub-2; and O, P, Q, R, S and T systems are independent because of different populations, except for the possibility of paired transposition. If this last difficult operation is performed the independent systems are reduced to twelve; but due to the systematic objections to rotation for simple structure in the transposed forms and the absence so far of a single successful example of transposition, it is safer at present to deal with twenty-four systems.

This is not the place for an exhaustive examination of possibilities of collation of observations for covariation study, which the student may take up elsewhere (**22**). But it may be pointed out that in addition to the six single strips possible in the six faces of the above chart,

there are also double strips in which an *increment or difference* in one dimension is used as the basis for correlation instead of an absolute value. For example, R-technique can be used correlating *increments* of ability after a lapse of time instead of abilities at a given moment. Or Q-technique can be used correlating the *differences* of persons, e.g., of twins, instead of their absolute measures; or P-technique can be used entering the correlations with the difference between two variables, e.g., speed and errors instead of a single variable measure, and so on.[1]

Some of these possibilities are experimentally impractical or require very careful thought to clarify the realm in which the discovered factors would lie and have meaning and predictive application. Most have yielded no harvest of experimental findings principally because no one has thought to apply them. Thus, for example, the only studies on the above R-technique design using *increments* of ability are two by Woodrow (**143**).

[1] There is also a whole new class of correlations which comes into existence when we add one or more extra dimensions to the covariation chart. These are considered in connection with new research designs in Chapter 20 and will not be systematically explored here, for they would merely complicate the understanding of the primary covariation chart. However, to meet the questions of the student whose curiosity may already stray into these possibilities, their existence may be briefly discussed. The extra dimensions arise principally through dividing the variable or test measurement into a stimulus condition and a response, and then from further subdividing these into independently alterable conditions and various aspects of the response. They arise only where there is experimental control.

For example, we might measure the response to an intelligence test in the usual way, while systematically varying (a) the intensity of motivation and (b) the temperature of the room in which the response is made. The occasions axis usually implies nonsystematic variation of conditions, but it would now split into two axes of systematic and independent variation. Thus, for example, a correlatable series could be made by holding test measurement and person constant and taking *two* conditions of motivation to be correlated with respect to a whole series of increasing temperatures.

The splitting of the response aspect of the test into several axes of covariation has already occurred in a factor analytic example provided by Hsü. He calls this an example of intrapersonal factorization to distinguish it from P-technique which is another of the possible designs—though a more basic one—in which only one person is involved. Here a series of emotional stimuli are presented to one person and the different *responses* or aspects of the response, namely, verbal, physiological, memory, etc. are simultaneously measured, thus creating as many series as there are aspects measured. From this either the response aspects or the distinct stimuli could be correlated, though Hsü advocates only one.

The essential point to keep in mind is that the "occasions" axis in the covariation chart, p. 109, is only incidentally a time series and primarily defines the *manipulatable or changeable aspects of the stimulus situation.*

"EFFICACY" OF FUNCTIONAL UNITIES

If we assert that g (general intelligence) or any of Thurstone's primary ability factors are true functional unities, this should be demonstrable not only in terms of individual differences but also in terms of growth. That is to say, if ability to solve arithmetic problems, to choose accurate synonyms, and to answer classification problems are simultaneously high in one person and low in another, so that we infer a unitary trait underlying them, they should also *increase* together through maturation, or exposure to a pattern of training, in the typical individual. A factorization of increments after the lapse of a year or more shows that some, at least, of the primary abilities exist as patterns of individual difference of growth.

Now it is possible that the application of the simple structure criterion to factorizations of the same variables by R- and P-techniques, or by other approaches to covariation suggested in the covariation chart, will yield recognizably the same factor patterns. But it is also possible that in one or another of the research approaches a factor present in one will be missing—systematically missing, regardless of sample—in the other. What are we then to conclude? The philosophers object to using the expression degrees of reality, since a thing either exists or does not exist. Consequently, in regard to the varying degrees of universality[2] that undoubtedly are (to our senses) demonstrable among factors, we may perhaps speak of *degrees of efficacy.*

A unitary trait may thus conceivably be unitary in some situations and not others, and the more efficacious factor is that which exercises its influence in more situations. For example, the parts of a cloud may move together, reflect light in the same way, and disappear together when heated by the sun, yet *mechanically* they have no unitariness, and a high degree of unitariness by other standards is also absent. For instance, one part may be cut off without affecting another. Or, turning to social data, we can demonstrate (**20**) a dimension of social status among occupations, loading prestige, earnings, average intelligence level, control of family size, etc. But no such dimension has yet been clearly demonstrated *among* as distinct from *within* nations

[2] This tendency of some factors to manifest themselves in many situations in which the variables are tried for correlation while others appear in relatively few reminds one of the geneticists' notion of differences in penetrance or expressiveness among genes or the chemists' notion of degrees of stability and range of incidence of various molecular structures.

when national groups of wide variety are compared (**27**), nor has it been shown that when an occupation undergoes a change of income level, all the other variables in this pattern change *pari passu*. In psychology there are countless R-technique proofs of the existence and the specific pattern of g or general ability, but so far this factor has never been demonstrated in P-technique studies—presumably because general intelligence does not fluctuate as a whole from day to day. Possibly, therefore, *every* functional unity breaks down or fails to show itself when subjected to some extreme circumstance or unusual test of covariation.

The discovery that a factor found by one covariational approach may be missing when approached through another situation or technique of covariation forces upon us, therefore, this notion of degrees of efficacy. The complementary discovery to this—that one and the same correlation matrix derived from one covariation study only may yield more than one simple structure solution and, therefore, *alternative sets of factors*—presents a more disturbing problem. Criteria will be discussed in Chapter 14 for deciding, when two or more apparent simple structures occur, which is technically the adequate one. But the fact remains that when such apparent instances have been cleared up, there still remain a very small minority of experiments where two or occasionally more alternative simple structure positions stand out quite clearly from the wide range of possible rotation positions. If technical statistical tests fail to demonstrate that one of these solutions is false, are there still other criteria which can tell us about the relative reality of the solutions?

EQUIVALENT SOLUTIONS

The outside checks mentioned in the last chapter may do so, but there are difficulties in leaving the verdict to these. There are times, seemingly, when we must face genuinely equal, equivalent solutions. This has sometimes been considered disconcerting to the factor analytic method, but it is not scientifically disconcerting to anyone familiar with the logic of unities or with modern physics (**87**). The two factorizations correspond to two alternative explanations of the same facts, as when we say that white light is produced either by adding the colors (wave lengths) of all the spectrum or by combining two complementary colors. Sometimes the alternative factors isolated by disputants will stand in a simple part-to-whole relation equivalent

to the alternative statements "John went to New York on the train" and "John went to New York in a Pullman coach." That is to say, one conceptualization breaks the unities down into coaches while another includes them in the larger unity of the train. Sometimes the difference of factorization will correspond to a difference of phase in the chain of causation, as when we say, "The mountains are not suitable for peach trees" meaning, "Low temperatures are inimical to peach trees." Sometimes we may have a genuine balance of efficacy between two factorial conceptualizations akin to the situation in a picture which may be seen in two perspectives or may be resolved into three or more alternative closures in terms of different objects. Thus many of the patterns of facts found in physicists' observations on radiation will fit equally a wave hypothesis or a discrete particle hypothesis (while perhaps some third explanation beyond either, equivalent to our second-order factor among factors discussed below, may be one solution here).

Embarrassing instances of real truth in alternatives occur most frequently in the social sciences where the absolute truth of each alternative is usually self-evident to one of the disputants. (These delicate situations are fair meat for persons with emotional prejudices!) Thus we may say that the increase of deaths from tuberculosis in World War II was caused by the war, by poor nutrition, by a tuberculosis epidemic, or by the toxins produced by germ action. The physiologist and the social scientist will talk in different terms here partly because they cut the chain of causation at different levels, as in the second explanation of factorial alternatives above, but also because of different, simultaneous, conceptual groupings of the same observations.

In the complex interactions of economics and history, it is easy to see that, at least up to a point, these alternative groupings are equally efficacious means of predicting and controlling the numerous variables, and that, as far as convenience is concerned, each can be preferred for particular purposes. Thus we can ascribe wars to the rascality of politicians, to international armament industries, to capitalism, to crusading communist imperialism, to high birth rates, to low intelligence of the masses, to frustration by the restrictions of civilization, etc. These are not entirely independent arguments; for example, the frustration by civilized restrictions is greater in the less intelligent and these are also prone to have high birth rates unrelated to op-

portunities. In its simplest form this particular balance of efficacy in formulations can be illustrated by the statements, "John was killed by the Cariocci family" or "John was killed by a South Side gang," where the gang and the family overlap in the members that did the deed. This phenomenon of alternative conceptualizations must not be confused with the fact of multiple causations alone, though multiple causation generally exists as a condition of alternative conceptualization.

It is not suggested that there are different truths here, but only that there are different natural cleavages in the data, each a true way of handling prediction or explanation problems and each of special convenience when one has particular data and particular predictions to handle. *All* these factor systems have to be taken into account in any adequate explanation. But for control and prediction it may be possible to work in one system providing one knows that the conditions are appropriate and that all the formulations remain consistently within the system. For example, when the student has understood what is meant by second-order factors, he will see that first- and second-order factors *regularly* provide such alternative factor systems, each order being adequate in itself to predict up to the degree of accuracy commensurate with its neglect of specifics.

SECOND-ORDER FACTORS

The notions of *degrees of efficacy* in factors and of *equivalence* in alternative conceptualizations are thus involved in and required as a background for discussing in proper philosophical perspective the technical problem of second-order factors upon which we must now enter to complete the student's survey of factor analytic concepts. Second-order factors arise only with oblique, correlated factors—a variety of factor which we must now describe. Although the demonstration rotation in Chapter 5 was carried out with orthogonal factors, i.e., reference axes kept at right angles to one another, it was pointed out that the rotated matrix often, indeed almost invariably, finishes up with the axes no longer at right angles. In almost all psychological, biological, or social correlations yet examined the nebulae of variables which fix the hyperplanes are themselves not exactly at right angles. The pursuit of true simple structure, with refusal to damage the simple structure for the sake of mere mathematical habits of tidy orthogonality, in general unmistakably directs the rotator to abandon

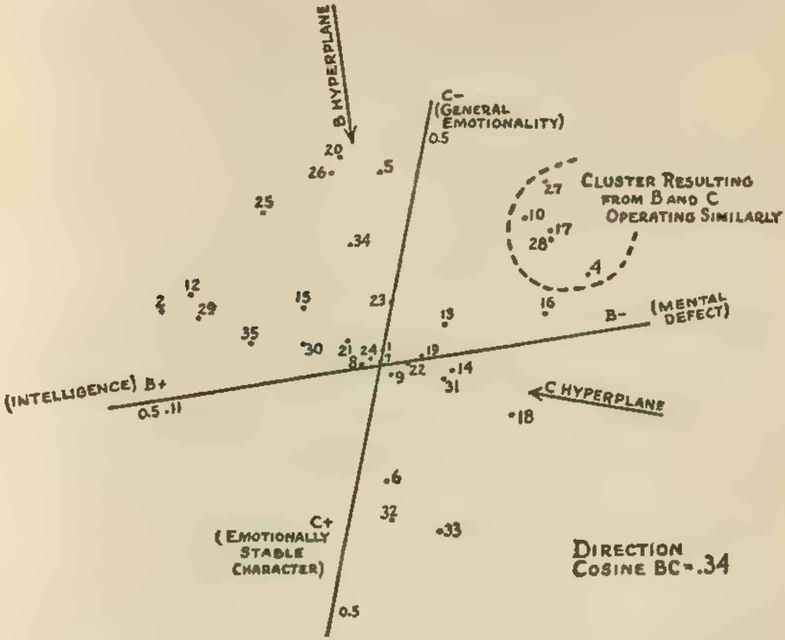


DIAGRAM 15. An Example from Psychology of Correlated, Oblique Factors. (From *Description and Measurement of Personality* by R. B. Cattell. Copyright 1946 by World Book Company. Reproduced by permission.)

absolutely orthogonal factors. All experience of rotation, alike with data of physical, biological, or social sciences, forces upon us the truth that in nature factors are correlated.

That factors, each of which acts as a powerful unitary influence upon a whole set of variables, should themselves be somewhat correlated is actually no matter for surprise. In a single universe everything to some degree influences everything; and though a factor may stand out as an independent influencer, as far as a host of variables dependent upon it is concerned, it is itself influenced by other factors. We deal with something analogous to the organization of society by the feudal system, in which a single organizer as viewed from the standpoint of his serfs is himself influenced by his peers and organized by his superiors.

As practical instances from psychology we may take Thurstone's correlated primary abilities or data from the personality realm where

it can be shown, for instance, that we tend to find a positive correlation between the reference vectors for factor B, *general intelligence* and factor C, *emotional stability* or ego strength, as shown by the hyperplanes in Diagram 15. (Incidentally this is taken from a selected population—students—and the selection may invert the usual correlation. There is reason to believe that in the general population, intelligence and emotional stability factors correlate positively about 0.3.) Although, as the hyperplanes show, these factors are separate organizers as far as most of the variables are concerned, they are themselves caused to correlate by some extraneous influence.

Incidentally this discussion of correlated factors will serve also to bring to our attention the two principal and different senses in which the term independent can be used. A thing can be independent in the sense that it can be independently analyzed out of the data, conceived, and labeled. In this sense a color, a shape, or number has an independent existence though in nature we always see a colored *object* (*or position*) or a number of *things* or *events*. Such independence is compatible with their having *mutual influence* as aspects of a single organism in nature, and though such attributes can be represented in mathematical equations by independent symbols, they tend also to be mathematically correlated. For example, white bears tend to be larger than brown bears, and the factor behind the coat color variables will, therefore, tend to be correlated with the factor behind the size variables in a general population. On the other hand we can have a more complete type of independence in which to independence of conception is added statistical independence of a more complete kind as when two variables are quite uncorrelated and are represented spatially by vectors at right angles. Thus color in bears might be quite unrelated to a factor of healthiness. The orthogonal factors which were sought in the early days of factor analysis belong to the second system, whereas the correlated factors which the exploration of nature has forced us to recognize as far more widespread have only the first kind of independence.

If factors are represented by vectors or axes no longer at right angles, it is obvious that we might conceivably find clusters among them, i.e., some cone of vectors such as we found among test vectors. Several such clusters can in fact be found among factors established in abilities and in the realm of personality, though they are *much more shallow*, i.e., have larger angles within them, than among test vectors.

In personality factors we find that emotional stability, intelligence, dominance, and surgency form one cluster while the three schizothyme factors form another, and so on.

It should be possible to take such a cluster and run a factor axis through it, or find nebulae among the factor vectors as we do with tests to provide a new hyperplane and erect a new *factor among factors* (73).

In such a search for *second-order factors*, as we call these factors among factors, the actual procedure would consist in taking enough sufficiently diverse test variables to give a minimum of, say, a dozen factors. One would then plot the diagrams of each factor with every other and make the best possible search for simple structure, seeking an unusually good definition of hyperplanes in order to fix the correct angles between the reference vectors. The cosines of these angles, i.e., the correlations, are entered in a vector correlation matrix just like a test correlation matrix. Actually this does not correspond exactly to the correlations among the *factors*. What is called the inverse of this matrix will have to be calculated, as explained in Chapter 13, to give the true factor correlation matrix; for what we get in our drawing are, technically, angles between reference vectors, not factors; but we need not digress into that transformation now. Suffice it that upon this factor correlation matrix obtained from the simple structure drawings we can now operate by the usual methods of factor extraction and rotation.

MEANING OF SECOND-ORDER FACTORS

The resulting factors are our *second-order factors*. The term superfactors has also been proposed, but is unsuitable because there are not just two categories of factors; there may conceivably be third- and higher-order factors to be considered. These second-order factors, if they are common factors, are naturally less numerous than the first-order factors which they organize. At present not enough instances have been explored for us to generalize inductively as to what their general nature is likely to be. Among personality factors a second-order factor is at present indicated running through the first cluster mentioned above. It is surmised that this is a factor of social status—an organization brought about among the factors of intelligence, emotional stability, surgency, and dominance by the fact that they are all

to some degree positively selected by upward social mobility. A more completely established second order factor is that which Thurstone finds, with some others, among the six or seven primary abilities. It is generally accepted that the most powerful second order factor here corresponds to Spearman's *g*, a general ability underlying the particular developments presented by the primary abilities.

From these instances, if we accept the view that Spearman's *g* corresponds to Lashley's concept of mass action of the whole cortex, we can see that the organization revealed in second-order factors may be produced either by the influences of the whole social structure upon the individual personality or by biological roots which are permissive for a particular group of psychological developments. However, on *deductive* grounds one would not expect all instances of second-order factors to point to organizers in outside realms, such realms as the neighboring sciences of sociology and physiology present in relation to psychology. The above preliminary inductions from the present sparse instances of second-order factors should not be allowed to lead us to any such rash generalization. Indeed, for the present it is unprofitable to speculate further about the general nature of second and higher order factors. It suffices to recognize that since all things are interrelated in our universe some degree of correlation would be expected, just as hyperplanes indicate, among our primary factors, and that we should expect to trace this correlation to further influences beyond our factors.

In actual fact the correlations so far found[3] among factors have generally been quite small, exceeding 0.4 in perhaps one case in ten and very rarely exceeding 0.6. As Thurstone has proved (**126**), the use of population samples that are strongly selected as to variance in some variables, while it will not affect the loading pattern of a factor, will modify the correlations *among* factors and therefore the loading pattern of the second-order factors. The definition and use of second-order factors is therefore likely to demand more precise thought about

[3] Incidentally one should distinguish between the true *r*'s found spatially among vectors by plotting their hyperplanes, as described above, and the *r*'s obtained from estimating the various factor endowments (by combinations of tests) of various individuals and then correlating the factor scores so obtained. Due to faulty factor estimation, particularly the use of tests for one factor that are also loaded in another, the *r*'s obtained in this latter fashion are spurious and err generally in the direction of being too high. (Witness the high correlations obtained between factors by Guilford on his questionnaire.)

concepts and fuller understanding of the meaning of calculations than are normally demanded of the employer of the simple specification equation.

COMPUTING IN FIRST- AND SECOND-ORDER REALMS

We must recognize that the use of correlated factors introduces complications of computation. Curiously enough these have caused mathematically rather than psychologically-minded practitioners to be slow in accepting them. In the first place the rotation to oblique positions may cause the correlated factors to occupy less *common* space than the space occupied by the original unrotated and uncorrelated factors. Thus if we have twelve factors before rotation, we shall also have twelve factors afterwards; but they do not so efficiently describe the space of the test vectors. They crowd into some parts of the space and neglect others and may leave a good deal of variance to factor specifics. In fact, the number of *second-order* factors which later emerge may be regarded as a measure of the loss of symmetrical occupation of the test space by the factors. Oblique factors also present complications in first-order calculation. The estimation of a single correlated factor is done as with an uncorrelated one by adding, with or without weights, the variables highly loaded in (not quite the same now as highly correlated with) the factor. But a complication comes in the specification equation which now alters its form, since for reasons that may become clearer in Chapter 13 the loadings (situational indexes) are no longer quite the same as the correlations of the tests with the factors. The formula for effecting the transformation from correlations to loadings is given in Chapter 13.

The specification equation for the second-order factors in terms of first order is naturally like the ordinary specification equation for first-order factors with respect to *their* variables. Thus if F_1 and F_2 are second-order factors and F_j is first order, we shall have:

$$F_j = s_{1j}F_1 + s_{2j}F_2 + s_jF_j \tag{17}$$

That is to say the variance for the population (or, in an individual instance, the score) on a first-order factor may be expressed in terms of the variance (or individual score) of second-order common factors and a second-order *specific* factor. Indeed all that holds for first-order factors in relation to tests holds for second-order factors in relation

to primaries.[4] For example, the second-order factors can be rotated for simple structure, and they may also finish in oblique positions. The chief difference is one of degree of accuracy of prediction. For most kinds of error of estimate accumulate, and, in addition, the second-order common factor loses control of the variance both of the specific factors in the tests and the specific factors in the factors. Consequently the definition of prediction from second-order factors is practicable only when experiment and calculation have been highly accurate from the beginning.

Not only can first-order factors be expressed by the specification equation using second orders, but the original tests or variables can also be expressed in the second-order realm, leapfrogging the first. Thus in a situation where one second-order factor appeared among four first-order factors, a performance P_j would be expressed as follows:

$$P_j = s_j F + s_{1j} F_1' + s_{2j} F_2' + s_{3j} F_3' + s_{4j} F_4' \tag{18}$$

where F is the second-order common factor, as before, and the F' factors corresponding to the unique part of them. The latter are the independent specific factor dimensions remaining after the second-order variance has been removed from the primaries. See Thomson **(120)**, page 297, and **(126)**.

NECESSITY OF OBLIQUE FACTORS

Our toleration of the inconveniences of oblique factors, as just recited in doleful survey, depends on our belief that in explaining and predicting natural events it is actually more convenient in the long run to follow nature than to attempt to force upon it some artificial oversimplification. Adhering to orthogonal factors gets us into more than complexity of calculation; it brings misleading errors, incompatibility of different research results, and downright contradiction. In the first place, since the angles among factors differ in different samples, the same loading patterns will never be reproducible by orthogonal factors in any pair of matrices. And those proofs of the functionally unitary character of the factor which may come from strategically placed experiments properly interrelated and from ulterior information (illustration on page 89) will not work. For the patterns

[4] Primaries is a term sometimes used to refer to the main first-order factors in any given field, e.g., primary abilities, primary personality factors.

from the various orthogonal solutions cannot be the same. An orthogonal factor can have meaning and stability only in its own matrix.

In short, *if we want to deal with the same factors in many different experimental, statistical, pure, and applied situations, with all the advantages of knowing the characteristics and natural history of the factor which that gives, we are compelled to deal with oblique factors.* As soon as economy is understood as *economy when considering science as a whole* instead of economy of calculation merely within a single correlation matrix, the criterion of simple structure and the resultant oblique factors are seen to present the better application of the scientific principle of parsimony for they give this overall economy. To require simplicity obsessionally in a single matrix alone is like an astronomer saying that he will recognize an object to be the sun only when it appears in his telescope as a circular disk of a standard size. This means that he will fail to recognize it when it is oval through horizon refraction or smaller through his being at aphelion, while conversely he will mistake an object for the sum which appears as a correct disk but which would not satisfy appearances that the sun should have in other contexts. And to insist on orthogonality of factors is indeed mistaking means for ends, since these simpler mathematical devices after all are only means to discovering and expressing whatever is in nature itself.

FACTORS AS EXPLORATORY "EMPIRICAL CONSTRUCTS"

The above debate on the meaning of unitariness and on possible alternative manifestations of organic patterns brings us back to a question about factor analysis in scientific method which we raised in the first chapter. It was pointed out there that factor analysis can be used as an exploratory device, a prelude to controlled experiment, aimed at discovering the variables corresponding to the proper unitary influences to be taken into account later in a controlled experiment. The truth of this can be more readily seen now that we have become acquainted with the phenomenon of correlated factors. If factors are not independent, then any two can be used as controlled and dependent variables in experiments to determine the more precise forms and laws of relationship between them, after a simple correlation coefficient has shown the existence of *some* relationship.

It has been pointed out that in this exploratory phase it is at first unimportant whether we do or do not start out with a *specific*

hypothesis as to the number or nature of the factors that exist. Indeed, in poorly controlled factorizations there is a certain danger in starting with a hypothesis, for it is to some extent possible to manipulate the number of factors (by communality assessments) and their nature (by rotation) to fit quite a range of hypotheses.[5]

Indeed it is best to use one factorization for hypothesis *production* and a second, distinct factorization for hypothesis *testing*. Thus after these first explorations the factorist will normally profit in his further experimental design if he is able to proceed with a fairly well-developed hypothesis. This hypothesis can spring from one of two levels of construction. First it may appear as an empirical construct, based and elaborated from the dimly perceived outlines of the factor as found in the pattern of variables in the initial exploration, as indicated above. But it may also appear as an ideal or logical construct, i.e., an idea obtained from reasoning by analogy or from borrowing in other ways a ready-made notion from some field remote from that of the actual data. For example, Franklin's notion of the cause of thunder was obtained by the empirical process of getting electric sparks from kites, and he understood thunder as a crackle of an electric spark on a larger scale. This empirical construct was correct but did not have much content whereas the ideal construct that thunder is due to the god Thor banging on his anvil has more ideational richness but, like many of our fine-sounding hypotheses, has very little contact with any facts.

Historically the tendency has usually been to attempt elaborate ideal constructs in the early stages of exploration. Only as man's mind becomes more disciplined, modest, and effective has it become more usual to be content initially with empirical constructs and to bring in wider, ideal constructs only when the area of knowledge is highly structured. Thus to take an instance from psychology we note that in the early investigation of rigidity there were many elaborate theories explaining the few factual connections noted (as well as many others not yet observed) in terms of neural physiology (Muller's secondary function) or equally far-reaching psychiatric generalizations. But the approach which eventually led to steady advance began with observa-

[5] As the reader will remember from the first chapter on scientific method, we are bound to start off with *some* hypothesis, e.g., that some order exists, among the variables. In a *second* factorization, however, the hypothesis becomes more precise and guides choice of variables.

tion of some restricted factual connections, indeed with the finding of a common factor through correlation of certain motor performance tests. The empirical construct put forward through examining the performances loaded in that factor was tentative in nature, but already very different from the ideal constructs in that it indicated rigidity to be essentially a resistance to change or learning, and on this basis research has progressed to a point where wider generalizations are beginning to be appropriate.

At the present stage of psychological knowledge, and perhaps of knowledge in the social sciences generally, much work with ideal constructs is liable to be presumptuous and fruitless. So much remains to be learned and so many simple connections remain to be established before any really involved reasoning has enough premises to go to work upon that the approach through extensive theories is inappropriate. Empirical constructs—small areas of law and order—need first to be formed, and factor analysis is powerful for this purpose. By this means the experimenter examines the crop of factors from the first exploration and inspects the highest loadings, which are initially likely to be only moderately high, in an attempt to infer what the essential nature of each factor may be. For example, he may find in psychological data a factor loading ability to answer riddles, ability to reconstruct hidden words, and ability to see hidden objects in a picture (**28**) or in sociological data for a population of countries a factor affecting frequency of political clashes with other countries, complexity of occupations, creativity in science and literature, death rate from suicide, expansiveness (gain in area) of the political unit concerned, etc. (**27**). By attempting to abstract what is common in each he may arrive at the hypothesis that the first is a factor of ideational inertia, and that the second is a dimension of *restriction of relatively direct instinctual expression* producing outward and inward aggression (suicide) as well as cultural productivity. He may then return with these hypotheses to experiment, using, along with the old variables, new variables designed to measure what his empirical construct defines. His hypothesis may then be confirmed to the extent that he is able to obtain higher loadings in the new, specially constructed variables than in the old. Indeed if his hypothesis is correct, the variables specifically introduced to represent the hypothesis should have loadings as near saturation as their reliability of measurement will permit.

In concluding Part I of this book, we may point out that it has been our intention to give a clear idea of the nature of factors, to bring the reader an acquaintance with the essential methods in their extraction, rotation, and use in prediction, and to indicate the role of the method in scientific research generally. This should suffice for the reader needing only to clarify his general ideas. But for the social scientist who is to use factor analytic methods in research or in actual applied predictions it is necessary to acquire greater confidence and facility by gaining in Part II experience of complete working methods. For only by so doing can he be sure that all implications of the general statements in Part I have been understood. The remaining chapters are therefore concerned with explaining the working methods in more detail and with facing technical difficulties which it was not necessary to discuss in giving a general introduction.

Questions and Exercises

1. Draw the covariation chart from memory and put in the pairs of series which represent the correlations in R-, Q-, O-, S-, T-, and P-techniques.
2. Discuss correlations of *increments* and indicate briefly the relation of the findings to other bases of correlation, in any area of science in which you know of results obtained simultaneously by this and other methods.
3. Compare intrapersonal correlation with P-technique and see if you can invent any further design on the lines of intrapersonal factorization utilizing the fact that the test can be split into situation and response.
4. What do we mean by the degree of efficacy of a factor pattern?
5. Give three examples, each if possible from a different area of science, in which two alternative sets of factors may be regarded as almost equally correct and useful in handling problems in the area concerned.
6. What is the primary argument for oblique factors and what secondary and general consideration support it? How are oblique factors estimated and why may the interfactor correlations from these estimates differ from those obtained by the first method?
7. Describe how second- and higher-order factors are obtained and give examples thereof. Set out and discuss the specification equation for (a) first-order factors and (b) single test variables, in terms of second-order factors.
8. Discuss the pros and cons for oblique and orthogonal factors.

Part II

SPECIFIC AIMS AND WORKING METHODS

CHAPTER 9

The Chief Alternative Designs in Factorizing a Matrix

It is our purpose in Part II to study methods of factor extraction and rotation in more detail and with more regard to acquiring efficiency in practical procedures. Thus we face certain technical difficulties passed over lightly in Part I. More general theoretical problems, however, will be left to Part III.

First the student needs to realize that the method of factor extraction advocated in Part I is by no means the only one possible or practiced. In demonstrating the method in Chapter 3 it was mentioned in passing that some radically different mathematical approaches can be made. There are, in fact, some five or six methods available having mathematically distinct goals. It is our purpose now to give the reader at least a nodding acquaintance with the objectives of these methods and the arguments regarding their relative suitability. In the following chapters we shall then go on to describe the more detailed working procedures specifically for the centroid method which has already been advocated and outlined and which is probably the method with the widest utility in the human sciences.

PRINCIPAL COMPONENTS METHOD

Of the factor analytic methods which work on a different principle from the centroid, probably the *principal components method* (also called principal factors and principal axes method) has the greatest appeal to mathematicians by reason of its elegance and determinateness. It may be read in detail in the writings of its inventor, Hotelling (**76**) or of Kelley (**80**), who has tried it on sociopsychological data and suggested various improvements.

If the reader will refer back to page 27, he will see that there

are two ways of representing a population of persons as points with respect to two test vectors—the usual correlation plot against rectangular coördinates, and the method of shifting the coördinates until the points distribute themselves about the origin with equal density in all directions. The principal components (or principal axes) method begins with the first type of plot in which a good correlation has been shown to be represented by a swarm of points taking an elliptical form. It can be readily seen that the usual correlation scattergram in two dimensions can be built up to three by adding a third variable along an axis at right angles to the other two. The elliptical swarm of points showing some degree of correlation between two variables will then become an ellipsoid—an egg—in three-dimensional space.

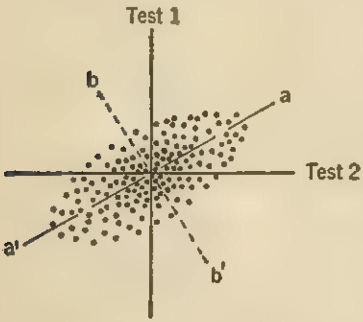


Fig. 1

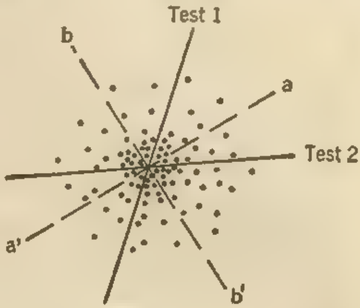


Fig. 2

DIAGRAM 16. Transformation of the Ellipsoid in Obtaining Principal Components.

And if we go on adding axes at right angles to the first three for all the variables in the matrix, we shall obtain a set of axes in imaginary multidimensional space—hyperspace—just as we do with factors in hyperspace. The student should note, however, that we are now using this hyperspace in a different fashion from the centroid method where test vectors were *not at right angles* but followed instead a convention of being at angles corresponding to their correlations.

Now the test vectors are at right angles and the individuals in the population form a swarm of points with an ellipsoidal form. In geometrical terms, the principal components method finds now what are called the principal axes of this ellipsoid and projects the scores of all individuals thereon. Thus in the two-dimensional problem used for illustration in Diagram 16, Fig. 1, the factors would be the princi-

pal axes *aa′* and *bb′*. With an ellipsoid of many dimensions one begins with the longest axis, takes the next longest at right angles to it, and so on through a diminishing series. After these axes are found, the scores (projections) on them are all brought to the same standard deviation, since by definition a factor is of unit variance—a dimension in its own right not to be compared with others. In geometrical terms this bringing of all axes to unit length means that the ellipsoid is squashed into a circular form, as in Diagram 16, Fig. 2. Since the people in their new positions still have to have their correct projections on the *test* vectors, as well as on the factors, the test vectors have to approach one another in the way we have already learned (page 43), as shown in Fig. 2. This results in the cosine between the tests becoming equal to r as in the centroid method and in the example already studied on page 36.

It will be seen that since the space is originally set up *to have as many dimensions as there are tests*, the principal components method ends also with as many common *factors* as there are tests. Such a lack of economy may scarcely seem to justify the labor of factor analysis, but we should consider the possibility (a) that the factors may make better psychological sense than the tests; (b) that the first few factors taken out actually account for most of the variance and in practice the rest can be neglected (for although all factors are by definition brought to unit variance this still does not enable the originally shorter factor axes to account for more than a minor fraction of the variance of the tests); and (c) that the number of factors is in any case a shade less than in the centroid method, for as we shall now see the latter actually has more factors than tests! The first of these propositions, as will be seen in a moment, remains only a possibility; in essence it is not true. The second is true. The third is technically true but requires some further discussion since we are accustomed to think of the centroid factors as being decidedly fewer than the variables they replace.

COMPARISON OF CENTROID AND PRINCIPAL COMPONENTS

Let us now take due note that in the centroid method it is only the *common* factors that are less numerous than the tests. Table 12, showing a complete set of specification equations comprising the factor matrix for an example with five tests and two common factors,

reveals that there are in all *seven* factors, *counting the specific factors,* viz: F_1, F_2, F_a, F_b, F_c, F_d, and F_e.

It will be seen that in respect to estimating the factors in the centroid method we thus have always more unknowns (seven here) than simultaneous equations (five here), so that the values of the factor endowments are, from a strictly mathematical point of view, indeterminate. That is to say, we cannot solve the equations and expect to get a unique set of values. For this reason the centroid factors are always only estimated with a margin of error, whereas the principal components, established from as many equations as there are unknowns (as many variables as there are factors), can actually be *calculated* with exact unambiguous solutions.

TABLE 12.

$$P_a = s_{a1}F_1 + s_{a2}F_2 + s_aF_a$$
$$P_b = s_{b1}F_1 + s_{b2}F_2 + s_bF_b$$
$$P_c = s_{c1}F_1 + s_{c2}F_2 + s_cF_c$$
$$P_d = s_{d1}F_1 + s_{d2}F_2 + s_dF_d$$
$$P_e = s_{e1}F_1 + s_{e2}F_2 + s_eF_e$$

Without space for further discussion [see, however (72)] the relative merits of the centroid and principal components methods may be summarized by saying: (a) the principal components method, by virtue of what has just been pointed out, permits completely accurate reproductions of the individuals' scores from the factors (not merely estimates), whereas the centroid permits only correct restoration of the correlations among tests; (b) the great fraction of the variance taken out in the first few principal components permits better calculation of scores or correlations than would be possible for the same number of first factors in the centroid extraction. This has no particular virtue since no one proposes to work with incomplete centroid factorizations; (c) the direction of the first principal axis, like the direction of the first centroid, is changed when one adds a test to the battery. For the first axis is, as it were, a sum of the directions of many variables, so that any new test adds variance which pulls the ellipsoid over in one direction or another. Consequently the principal components have no fixed psychological meaning and are *at the mercy of the particular choice of tests in the battery.* To arrive at any unique psychological or scientific meaning, they need, like the centroid fac-

tors, to be rotated later for simple structure. As stated previously, all the factorial methods can have their results transformed into one another or into some common end result, e.g., simple structure, and if this latter aim is admittedly the correct one, the principal component method has no virtue in itself and no claim to preference unless its computing processes are shorter than the centroid calculation. This is not the case; the calculations are if anything longer and more difficult. Consequently, if the student is to learn one method only, he had best learn the centroid method, though this digression into principal components is desirable to give him enlightenment on the relation of the methods.

BIFACTOR METHOD

Another method of factorization which deserves our attention is the bifactor method (not to be confused with the bipolar method). For the student of factor analysis the method of two-factor analysis, invented by Spearman, has the historical, genetic interest that the Rocket locomotive or the Wright plane has for railroads and airways, respectively. But this older approach, rooted in the special problems of intelligence measurement, has been developed into a more generally applicable process called bifactor analysis by Holzinger (**71**). The reader may recall (page 49) that Spearman used the tetrad difference equation, or the possibility of arranging *r*'s in a hierarchy in a matrix, to prove that the tests in the given battery could be explained by (1) a single general factor running through all, (2) a factor specific to each, and (3) no other factors whatever. If any test broke the hierarchy and upset the tetrad relation by bringing in a second common factor, Spearman threw it out of the battery, for he was interested only in measuring the general factor—the intelligence factor—in cognitive tests. Other psychologists, however, were more interested in what went into Spearman's waste paper basket than in what went into the test battery. They gathered the outlaw tests which broke the hierarchy to see what number and kind of group factors might hold some of them together in addition to the general factor which they shared with the more well-behaved tests (**116, 133**).

The procedure for finding these group factors begins essentially with the Spearman method of estimating the amount of general factor

in the tests which do *not* break the hierarchy. The correlation among tests which *do* break the hierarchy is then worked out in so far as it is due to the general factor, and this value is subtracted from the observed correlations. Since, when all correlations are positive, the general factor first taken out is, as it were, the least common denominator of all the tests, the operation of other factors that bind certain restricted groups of tests together is easier to see if they also are all positive in nature, i.e., if all the residuals are positive in the amount of correlation that remains to be accounted for *additional* to that due to the general factor. The group factors then appear as nonoverlapping growths on top of the general factor. The system is akin to describing the Rocky Mountains by a basic altitude running under all the separate mountain ranges and then adding a separate statement about the additional altitude of each separate range. For reasons which will be clearer later, this bifactor pattern of analysis has also been called the hollow staircase pattern.

The computations for the Spearman general factor loadings and for bifactors, which are not as complex as those for principal components or even for the centroid method, can be read elsewhere (**11, 71, 74**). An extension of this method is called "factor analysis by submatrices" (**10**).

ARTIFICIALITY OF "GENERAL" AND GROUP CLASSIFICATION

In passing, we may note that the distinction between general factor and group factor which was made much of in the literature at the height of the Spearman researches and in early debates on factor analysis has had restricted utility since multifactor analysis developed. The factor which appeared as an interfering group factor in a Spearman intelligence hierarchy, e.g., a factor of verbal ability or perseveration or fluency, often turned out to be the important general factor in another battery. What is general to a battery of variables obviously depends on the choice of variables in making it up, and this choice generally followed the experimenter's ideas of what was most important or interesting. Consequently general is a relative and subjective, not a universally valid, term.

Spearman's method certainly had its feet firmly on the ground of first principles and avoided some of the more upsetting possibilities of error which arose with the bold flight into hyperspace undertaken

by multifactor analysis. But, in congratulating themselves on this firm foundation the workers with general and bifactor methods have been prone to overlook that the *position* of the firm ground on which they stand is not necessarily the center of the universe, and that it is indeed highly relative and questionable. From the perspective of intelligence testing many things could be called group factors which, in regard to the total personality, turned out later to be quite as general as intelligence.

Indeed what is general or group or specific, in terms of a correlation matrix, depends intimately on what has been chosen to go into the matrix. If we chose only intelligence tests, intelligence will certainly be a general factor. If the experimenter has some *real* basis for claiming that he has an even, stratified sample of all possible psychological tests, i.e., a basis having a wider reference than the merely arbitrary correlation matrix of a given research, then general, group, and specific may have some meaning; and this basis is perhaps provided by the *personality sphere* concept (**22**). But by such a touchstone it turns out that *no* factor is found to be truly general or omnipresent. Every factor is found to have a hyperplane of negligible or zero (apart from chance) loadings. The notion that every source trait affects *every* act and aspect of personality is true in principle, but it is also true that one can distinguish a set of performance situations in which the influence is powerful from a host of others in which the influence is so slight as to put the variables essentially in a hyperplane class.

The distinction between group and general factors therefore loses any meaning—except in a particular matrix and in the particular sense used in bifactors—and it is better to call both of these older concepts *common factors of greater or lesser coverage.* The specific factor still remains, but for reasons given in Chapter 18, specifics are probably much less important in the conceptual scheme of things than current thought supposes them to be. Actually they can generally be brought readily under the heading of a narrow *common* factor by multiplying tests very similar in nature to the variable having the specific. For example, a test of sorting wool colors may have a large specific factor when placed among general ability tests; but if we invent a dozen or so tests all involving some skill with colors and shades, it is likely that this specific variance will largely become

shared with these other tests. In short it is likely that, in the end, general, group, and specific factors can be brought under the common concept of common factors.

BIPOLAR FACTOR SYSTEM

A method of analyzing correlations which is historically related to the Spearman general factor and Holzinger bifactor approach is the bipolar factor system of Burt. It is actually a general factor method and does not necessarily require breaking down the matrix into submatrices for grouping as Holzinger's method does, but it is on the same footing as Holzinger's method in that it accepts the first general factor taken out as the foundation which determines the shape of later factors and treats it as a *real, final factor* not to be interfered with or divided up by rotation. In this method the first factor, which is as usual an average for all the correlations, has wholly positive loadings (if eccentric negative tests have been reflected). But the later factors do not have positive loadings only, as in the bifactor method, because the first factor takes out more variance than in the latter method. In fact the second factor has negative loadings on those tests which have negative residuals after the correlation due to the first factor has been removed and positive for those whose correlations were above the average correlation in the first factor matrix. The second factor thus balances about zero with about as many tests positively as negatively loaded.

By the very nature of the extraction process, subsequent factors tend to arrange their loadings in a peculiar pattern such that each factor makes half the variables positive and half negative among those that were all of one sign in the preceding factor. For easy designation the present writer has suggested the term *genealogical* for this tracery, since it suggests the split into a male and female line in stepping back from generation to generation. Actually this method of factorization is in essence the centroid method; but it sets out to preserve the peculiarities of the loadings as they come fresh from the extraction process, and it groups the tests in blocks according to the signs of the loadings, whereas in the main centroid method which we set out here these peculiarities are disregarded and, in any case, soon become lost in the rotation process. Besides, in the shorter methods of computation (grouping methods) into which the centroid method becomes improved, the factors lose all relation to Burt's genealogical bipolar factor pattern even at the stage of factor extraction itself.

CONSTELLATION, STRUCTURE, PATTERN, CONFIGURATION AND RESOLUTION

Looking back over the four methods described (centroid, principal axes, bifactor, and bipolar factor), the reader will see that they differ in the mathematical devices and concepts used in factor extraction (except in the first and last), particularly in the case of the principal factor solution. But since, as Burt has pointed out (**11**), these various solutions are all capable of transformation (e.g., by rotation) one into another and (except for some features of principal factors) are mathematically equivalent, the psychologist's choice among them and their inventor's claims for preferring them must depend upon considerations beyond mathematics. Essentially the preference tends to depend upon either (a) the fact that the particular rotation position, i.e., the particular constellation of factor loadings,[1] given by the process immediately without further rotation, is in some way especially scientifically meaningful and useful or (b) the fact that the computation processes are simpler or quicker.

Let us examine first the question of whether certain general patterns or constellations of loadings have special virtues. The possible loading constellations have been set out with particular attention to detail by Holzinger and Harman (**71**, and by Burt **11**). One must dis-

[1] The terms constellation, configuration, pattern, and structure have come to have quite specific meanings in factor analysis which must at this point be clarified as far as our stage of exposition permits. Unfortunately different authorities, notably Holzinger and Thurstone in respect to the term structure, have adopted different meanings, which fact presents us with a task of reconciliation.

By a *constellation* we shall mean the general arrangement of the loadings among factors with respect to any systematic plan of zero, positive, and negative loadings as discussed in these paragraphs.

For this we might be inclined to use factor *pattern* if it had not already been definitely preëmpted to mean the factor matrix or table setting out *the actual loadings of the variables in the factors.* The constellation thus refers to the *general* characters of the pattern.

A factor *structure* is defined by Holzinger as the corresponding table of *correlations* between the variables and the factors. These are normally the same as the loadings, but not in the case of oblique factors as we shall see later.

Thurstone uses *configuration* to refer to the positions and relations of the test vectors in space—a system fixed by, and representing, the correlation matrix itself. When this configuration has *any* system of coördinate axes imposed upon it, it is said to constitute a *structure* and in one particular position a *simple structure.* Since Holzinger's distinction between pattern and structure is useful in discussing oblique axes, it would be clearer to call Thurstone's structure a factor *resolution,* i.e., a particular resolution of the configuration into factors, and this usage we shall follow.

tinguish, however, between the designs which might conceivably be obtained and those which the facts of science and laws of statistics actually make possible. The former are arrived at by setting down all the theoretically possible combinations of general factors, overlapping group factors, nonoverlapping group factors, and specific factors. The most important—or at least the most mooted and discussed—possibilities are set out schematically in Table 13 which illustrates factors for an example of ten tests and represents every significant loading by a ″ sign.

Figs. 2, 4, 5, 6, and 7 represent factor constellations that can be obtained with any normal correlation matrix, though only four of these have been advocated as having special merit. Figs. 1 and 3 represent constellations obtainable only with special conditions among the correlations and Fig. 1 has been advocated as psychologically meaningful. Figs. 3 and 4 thus represent those conceivable but not advocated constellations of which there are many more mentioned above to be possible from different *a priori* combinations of general, group, and specific factors and various sign patterns. A recent instance

TABLE 13. Possible Constellations of Factor Structure[2]

		Specifics									
Test	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}
1	+″	+″									
2	+″		+″								
3	+″			+″							
4	+″				+″						
5	+″					+″					
6	+″						+″				
7	+″							+″			
8	+″								+″		
9	+″									+″	
10	+″										+″

One general and specifics:
Spearman's hierarchy
(By Spearman's method all loadings were positive, but any pair of loadings, i.e., a general and a specific, can be made negative by reflecting a variable.)

Fig. 1

[2] In these drawings the symbol ″ represents the presence of a numerical loading. A positive or negative sign is attached to it only if the loading is compelled to have a particular sign.

Test	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}	F_{13}	F_{14}	F_{15}
1	+″	+″					+″								
2	+″	+″						+″							
3	+″		+″						+″						
4	+″		+″							+″					
5	+″			+″							+″				
6	+″			+″ ← Hollow staircase								+″			
7	+″			+″									+″		
8	+″				+″									+″	
9	+″				+″										+″
10	+″					+″									

One general factor and nonoverlapping group factors
Holzinger's bifactor analysis.
(Sign possibilities as with Spearman)

Fig. 2

Test	F_1	F_2	F_3
1	″		
2	″		
3	″		
4	″		
5		″	
6		″	
7		″	
8			″
9			″
10			″

Nonoverlapping groups (with or without specifics)
Holzinger's unifactors. (Any signs)

Fig. 3

Test	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}	F_{13}	F_{14}
1	″	″		″	″									
2	″	″		″		″								
3	″	″	″				″							
4	″	″	″					″						
5	″		″	″					″					
6	″		″	″						″				
7	″		″	″							″			
8	″			″								″		
9	″			″									″	
10	″			″										″

One general, overlapping groups and specifics. (Any signs)

Fig. 4

Test	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}
1	″	″	″	″	″	″	″	″	″	″
2	″	″	″	″	″	″	″	″	″	″
3	″	″	″	″	″	″	″	″	″	″
4	″	″	″	″	″	″	″	″	″	″
5	″	″	″	″	″	″	″	″	″	″
6	″	″	″	″	″	″	″	″	″	″
7	″	″	″	″	″	″	″	″	″	″
8	″	″	″	″	″	″	″	″	″	″
9	″	″	″	″	″	″	″	″	″	″
10	″	″	″	″	″	″	″	″	″	″

As many general factors as tests
Hotelling's and Kelley's principal factors.
(Irregular positive and negative signs on each factor)

Fig. 5

Test	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}	F_{13}	F_{14}
1	+″	−″	+″	−″	″									
2	+″	−″	+″	−″		″								
3	+″	−″	+″	+″			″							
4	+″	−″	−″	−″				″						
5	+″	−″	−″	+″					″					
6	+″	+″	+″	−″						″				
7	+″	+″	+″	+″							″			
8	+″	+″	−″	−″								″		
9	+″	+″	−″	+″									″	
10	+″	+″	−″	+″										″

General factors with genealogical sign patterns (plus specifics)
Burt's bipolar factors.

Fig. 6

Test	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}	F_{13}	F_{14}
1	″			″	″									
2	″	″		″		″								
3	″	″	″				″							
4	″	″	″					″						
5	″		″						″					
6										″				
7		″	″	″							″			
8		″		″								″		
9				″									″	
10			″	″										″

Overlapping group (in principle, *general*) factor and specifics (any signs)
Thurstone's multifactor method with simple structure.

Fig. 7

of such further possibilities is Guttman's proposal to analyze in terms of a chain—or even a cyclic order, i.e., a circular chain—of overlapping common group factors.

CONSIDERATIONS IN CHOICE OF METHOD

With this experience of a number of possible solutions, the reader may seek some method of classifying them as a preliminary to evaluating their merits for various purposes. A number of useful principles suggest themselves for divisions as follows:

1. According to the *constellation of factor loadings intended.* By this principle we can divide factorization methods into those which analyze the whole matrix at once, such as principal factor and multiple factor methods, producing one or more *general factors* to be contrasted with those which divide the matrix into subgroups, such as the bifactor or certain submatrix methods producing definite, *non-overlapping group factors.*[8]

2. According to some one basic characteristic of the *algebraic process* by which loadings are obtained. As shown in Chapter 3, the centroid method simply adds the column of correlations to get the *mean* r (to be exact, the *mean* r divided by the centroid r) of the given test with all others. This is its loading. By contrast the algebraic process for the principal factor solution (which we have not described in detail) gives a *weighted* sum of each column—weighted by a process of successive approximations—as the factor loading. Of the possible classifications according to the algebraic process, the division into processes of *simple summation* and *weighted summation* (**11**) is one which is fundamental for the mathematicians and which incidentally accounts for the greater determinateness of the principal axes solution. It is also an important consideration for the computer, the weighted summation methods being longer and more cumbersome.

[8] From the point of view of the general scientific interpretation to which they lead, the principal component and bipolar constellations, though different in computation, are essentially similar. Both take out as much as possible in the first factor and have bipolar factors remaining. The first component happens not to be all positive in loadings, as Burt's first factor is in his bipolar system, but this can be arranged by arbitrary reflections, as is done in Burt's first factor. The chief difference then remaining between them is that the first *component* is a *weighted* sum. There are also mathematical differences in the total plan—in that the principal component system defines the individual scores completely and has no specific factors. But in comparison with all other systems the general form of their constellation of factors vis-a-vis variables is similar.

3. According to the number of factors required to explain the tests. Obviously the design of Figs. 1, 2, and 3 are more economical than the others in this respect if we consider the total number of factors required, regardless of whether they are group, general, or specific.

4. According to whether it is proposed to rotate the solution once it is obtained. Burt claims that the bifactor and bipolar methods have the advantage of giving immediate psychological meaning without rotation, but this we question.

5. According to ease of computation.

6. According to completeness and accuracy of the determination of numerical values, e.g., of the factor estimates or the factor loadings.

There are further bases for classification and choice, but these are generally considered the most important and will alone be mentioned in whatever further discussion we give to the suitability of various factorization methods for various purposes.

It will be recalled that in asking the reader at the beginning of Part 2 to broaden his field of vision and survey the alternative possible designs in factorization, before settling down to one, we eventually suggested that two major criteria should direct the choice: scientific meaningfulness of the result and ease, accuracy, etc. in computation. Although these pages may have given the student a nodding acquaintance with the various designs, sufficient for a general orientation, a more detailed understanding would be necessary for him to follow the pros and cons for the various methods with complete insight. Consequently the necessarily brief arguments now to be added may unavoidably seem dogmatic.

Burt's arguments for the bipolar solution are prefaced (**11**) by a discussion of the particular relation (beyond the usual transformation possibility general to all methods) of bipolar to bifactor solutions which is of some interest to the researcher likely to encounter analyses in both of these forms. Incidentally, Burt regards the division into group and general factor methods (with respect to a particular matrix) more seriously than we do here and considers the bifactor and bipolar factor solutions as two most important examples respectively of group and general factor solutions. He then proceeds to show that the bipolar factors are *difference* factors with respect to the bifactors. Thus if the bifactors (after the general factor is accounted for) are verbal ability, numerical ability, spatial ability, etc., then the first bi-

polar factor (after the general factor) will be one measuring verbal ability versus the other abilities, i.e., nonverbal abilities. The bifactor will have positive loadings in verbal ability tests and zero loadings in the rest; the bipolar factor will have positive loadings in the verbal tests and negative loadings in the rest.

In other words the bipolar factors will tend to express ratios or differences where the bifactors will express absolute measures. For example, Burt presents a study of body-build factors in which, after the general size factor is taken out, the first *bifactor* is a head-size factor and the second, a trunk-size factor. The first *bipolar* factor, in the same data, loaded variables which showed it to be expressive of *head-size relative to the rest of the body*. In describing our perceptions, we often use adjectives which express immediately such relations in the total Gestalt, e.g., a "round face," corresponding to bipolar factor descriptions, but in most scientific work data can be more effectively handled by dealing with the simple absolutes e.g., a face *so* high and *so b*road (which can always be converted into ratios if we wish) rather than with the relations.[4]

The mathematical relationship between the bifactor solution and the rotated centroid multifactor solution can also be readily expressed in verbal form. Burt argues that in meaning the factors are essentially the same—since both have simple structure in the sense that every factor has zero loadings with respect to many variables. He therefore suggests that they have essentially the same psychological meaning and that the supposedly shorter computation form of the bifactor method should be adopted instead of a centroid analysis followed by a lengthy rotation (see also Swineford **118**). The objection to this is that the bifactor method is not necessarily shorter—especially with many variables, for it involves rearranging the order of the variables in the matrix in such a way that they fall in groups of submatrices corresponding to the anticipated bifactors. With numerous

[4] This avoidance of factors which represent relations of parts is not to be adopted as a rigid rule. In defense of its adoption, one writer has pointed out that one can conceive a growth hormone which makes for length of trunk and another responsible for growth of the long bones, but not one responsible for their relative growth. This example is perhaps a better contribution to the debate than to its author's particular argument because hormones are known which *do* suppress growth in one part and stimulate it in another! There is indeed no reason why a factor should *not* take bipolar form, and in the multifactor simple structure analysis they are at least free to do so. But there *is* good reason for objecting to *all* factors being forced to this form as in the general bipolar method.

variables and many factors this is laborious and beset by possibilities of error.

However, the real objection to Burt's argument for bifactor methods is that the factors obtained are *not* exactly the same in meaning as those obtained from simple structure centroid analysis. They are only approximately the same. In practice most of the factors may at first look very much the same, i.e., may have the same variables high in them. For example, the extensive work in Britain on analysis of abilities by Vernon (**133**), Stephenson (**116**), Burt (**11**), and others in which verbal, numerical, practical (K factor) and other special abilities have been brought out as group factors after the extraction of a general intelligence factor along Spearman lines, shows excellent apparent correspondence with the results of multifactor analysis in America based on rotations of centroid extractions which leave no powerful general factor. And the use of the strict bifactor method yields mainly the same groupings of abilities as have been found in the Thurstone primary abilities based on centroid rotations, except that the former method has not penetrated so well into the factor space and has still not yielded *all* Thurstone's primaries.

But certain theoretical differences remain which will not be without practical consequences in the last resort. In the first place the bifactor method sets aside a substantial general factor and the multifactor does not. In the second, the special ability factors are orthogonal in the bifactor and oblique in the multifactor presentation. Therefore the bifactor solution corresponds less to the multifactor pattern than it does to the arrangement of orthogonal second-order general factors and primary specifics obtained by factorizing the primary (multi) factors themselves (**65, 120, 121**), (equation (18), page 122). Both of these differences turn upon the fact overlooked in the above comments on similarity of factors, that there is *one* factor in the bifactor series of factors, namely the first general factor which is *never* matched by anything in the multifactor series. And on closer examination the remaining bifactor factors, which have generally lesser variance than have the multifactors which roughly match them, will be found to match loadings rather poorly. The agreement appears only in the really high variables, the loading profiles being more divergent for the variables with moderate loadings.

For some purposes there is perhaps little to choose between these two formulations, namely, the rotated centroid, yielding correlated

primaries, or the bifactor (or second-order multifactor) analysis yielding a large general factor plus what is left of primaries. Where this statement holds, it holds regardless of whether we are concerned to describe a person or to predict a performance. It fits our usual ways of thinking just as well to describe a person as having a certain level of general ability plus certain levels in special additional areas of ability as to define his level in various primary abilities, knowing that they are positively correlated. But the bifactor solution (as well as the bipolar factor solution for that matter) have this cardinal weakness when compared with the multifactor solution: that the first large general factor which is extracted is not invariant, i.e., not constant in loading pattern and therefore in meaning.

Since the score on the first factor is essentially the sum of the scores on all the tests in the battery, the addition of any new test affects the whole meaning of the general factor.[5] Its meaning depends on the sample of tests used and can be pulled over in any direction like a democratic vote by stacking the polls with test members from a particular area. This is true also of the first centroid factor in multifactor analysis—but in this case the later rotation does not allow the factor to remain where it grew up. There the rotation for simple structure decides eventually where the factors shall rest and experience has shown that *factor patterns obtained by simple structure tend to be invariant* (**29, 77, 97, 120, 126**). That is to say, the loadings of a set of variables form into the same factor patterns regardless of what additional variables are thrown into their company. Consequently that design seems safer and more likely to yield a configuration corresponding to whatever real functional groupings exist in nature, which rotates to the simple structural multigroup factor position instead of accepting the position given immediately by the bifactor solution, even though more work is sometimes involved in the former.

Of the six ways listed above in which factorial systems may differ one from another, and by which the merits of any given system may be assessed, the most important is undoubtedly the first, namely, the form of the factor constellation. Since unitary influences in nature may undoubtedly operate in a great variety of ways, sometimes over-

[5] However if the first general factor is extracted as in Spearman's method, additions *do* leave it invariant, *providing any added test falls into the hierarchy formed with those already in the battery.* In a chance selection of tests the addition of a new test would only rarely meet this special condition.

lapping in their influences, sometimes discrete, sometimes affecting all variables, sometimes affecting only a few, the only constellation plan that is acceptable is one that is *flexible* to reality and capable of permitting the emergence of whatever natural structure exists in the data. The only method which will meet this requirement is one using rotation and, indeed, the unrotated bifactor and bipolar patterns as well as the principal axes solution as first obtained are ruled out at once. The most likely configuration of factors is one of overlapping group factors (in principle, general factors with negligible loadings in many variables) with any combination of signs that may be required. Systems which restrict themselves less flexibly to one particular configuration of factors must be rejected, despite any claim to greater mathematical tidiness or ease of computation, because they require constellations, such as nonoverlapping influences, which are contrary to all we know about the interaction of psychological and social forces. These other methods also fail because they have not proved themselves capable of giving invariance, i.e., of reproducing the same factor patterns in different experiments and matrices. At best they have some use as quick classificatory and descriptive devices when employed upon tests in a single matrix.

Choice on the basis of ease, or alternatively, of minute accuracy, of computation is ridiculous if the computation does not give what is scientifically meaningful! Bifactor methods have claims for brevity of computation, and principal axes for accuracy of prediction of the individual performance. But when we talk of economy of calculation, it is necessary to have in mind that machine aids may soon alter the picture. Though one may form at present general impressions of the relative economy of computations by various methods, e.g., that the bifactor is quicker than the centroid and the centroid is quicker than the principal axes, and this in turn than Lawley's maximum likelihood method referred to below (page 396) and so on, the range of machine aids available and the extent to which they can be fitted to the various methods leaves the whole question of relative time wide open to a special research inquiry. For example, in analysis by submatrices the actual additions of correlations, etc. within clusters is admittedly decidedly quicker than handling the whole matrix by the simple centroid method. But first the researcher has to group the variables in clusters and reconstruct the matrix accordingly, while

the reflection problem is a considerable chore when the matrix does not consist wholly of positive correlations.

In short, though it may be worth while when research settles down to three or four methods to do a "cost accounting" study on the time and cost of various processes, a choice according to computing ease is at present impossible, by reason of our ignorance. Such a cost accounting will need to take into account (a) the level of accuracy generally desired; research to discover the nature of factors, for example, can be carried to fewer decimal places than that required for specification equation and predictive work; (b) the nature of the checking processes essential to the various methods; (c) the parts that can be fitted to electric computing aids, I.B.M. methods, and electronic computers; for example, the time for rotation becomes greatly reduced with electronic matrix multiplying computers; and (d) the levels of skilled assistance required and the extent to which the skills of the craft possessed by key workers; e.g., those practiced in estimating communalities and judging rotations, can shorten the process.

Providing rotation is to be carried out, any of the methods can be used to obtain the unrotated matrix and since we cannot at present choose among them on grounds of ease of computation (for the same given accuracy of end result), we are left to decide among them, as far as computational process is concerned, upon the remaining principles above—those distinguishing the methods according to degrees of accuracy and determinateness.

In regard to determinateness of the factor loadings and the prediction of individual performances, the centroid, bifactor, and bipolar factors fall in the less accurate groups, while the principal axes and the maximum likelihood fall in the more accurate and determinate solutions. The latter method, not yet encountered, must be briefly described. It begins by getting a first approximation to the factor loadings by the centroid or some other method. It then operates upon these loadings by a successive approximation method applying a correction each time the process is repeated. (This is not the same as the successive approximation process with the centroid using continually better communality estimates.) At the same time it employs a definite hypothesis as to the number of common factors that are really involved. With a study based on an adequate sample, a definite χ^2 test can be applied to see whether the first hypothesis tried is

correct, and so one can eventually arrive at a true estimate of the number and nature of the factors.

The Lawley method, which may be most succinctly read in Thomson (**120**) and in more detail in Lawley (**83**), is a very long process when there are many tests and factors; and whatever computing devices are used, this and the prolonged weighted summation methods, *e.g.*, principal factors, are decidedly longer than the centroid (without repeated estimation of communalities). At the present stage of most applications of factor analysis to social science where it is a question of discerning the main functional unities rather than of catching the faintest factor or making very accurate individual factor estimates and predictions, the writer believes that the centroid method of extraction is unquestionably the best and the most universally useful.

When the main factors have been mapped and the variances of various populations are accurately known so that the main influences in and conditions of predictions in various populations are known, the time will become ripe to calculate the unrotated factor matrices by the more accurate methods just described. Still, rotation to the general constellation of a multifactor (overlapping group factor) solution will remain essential for completion of the analysis. Meanwhile the best general-purpose method of initial factor extraction is the centroid and its immediate derivations. Upon the more finished developments of this method we shall now concentrate our attention in the next and some later chapters which are devoted to describing in practical form the best routine procedures in factor analysis.

Questions and Exercises

1. Explain how the centroid method of factorization always produces more factors than the number of tests used.
2. Compare the centroid and principal axes methods of factorization as to number and accuracy of factors produced, ease of factor extraction, and reliability of interpretation of the results.
3. Define what is meant by bifactor analysis, general factor, specific factor, group factor, configuration, constellation, and factor resolution.
4. Determine which tests in each figure of Table 13 have general factors, which have specific factors, which have group factors, and which have nonoverlapping group factors.
5. List and describe six ways in which the various systems of factorization may differ one from another and indicate which you consider most important for deciding on the ultimate usefulness of a system.

6. Explain what are meant by bipolar factor and genealogical constellation patterns and explain with examples the ways in which the results of bipolar, bifactor, and multifactor analyses resemble and differ from one another.
7. What is the essential weakness of those methods which first take out a general factor and then take no steps to rotate it into simple structure positions, and why does this weakness sometimes not exist in a Spearman general factor analysis?
8. Discuss the resemblance of a bifactor solution to the second-order factor resolution duly rotated to simple structure.

CHAPTER **10**

Working Methods for Centroid Extraction Including Communality Estimation

From the general survey of the relations of the centroid and alternative analysis systems carried out in Chapter 9 we shall settle down in this chapter to fuller investigation of the centroid method as being the statistical tool best adapted to most scientific uses of factor analysis. Indeed, since the limitations of the student's time make it possible to give him a real working familiarity with only one method, the remainder of the practical computing discussion in this book will center upon centroid extraction, and rotation for simple structure, though the generalized discussions which follow in Part III apply to simple structure factors obtained by no matter what computing devices.

THE BASIC CENTROID METHOD

A fuller historical account of the centroid methods, together with a more systematic and mathematically elegant presentation of working methods than is possible in this introduction will be found in Thurstone's *Multiple Factor Analysis* (**126**) which the advanced student should read in due course. Our purpose here is to give the reader sufficient working instructions to enable him to compute intelligently according to three principal methods—the full-dress *centroid method*, the shortened but still very simple methods known as the *group and grouping methods* and the still more shortened but more complex method known as the *multigroup method* of factor extraction. In this chapter only the first and most basic method will be described, together with some discussion of the problem of choosing communalities and other questions common to all three of the centroid-derived methods.

The *essential* steps in the basic centroid method have already been described in Chapters 4 and 5, so that it remains here to deal only with certain practical refinements, computational checks, and minor theoretical issues not described in that general and elementary approach. In the first place the *initial r* matrix is usually not wholly positive as in our prepared example, i.e., the columns do not all add to positive totals. Positive matrices are found only in special fields, e.g., general ability batteries where selection of tests, and care to score them in the right direction, provide this condition. Our example (page 41) resembles many of the historical instances in the ability field and was chosen from this restricted field merely to simplify initial exposition. In general, therefore, before starting on the first addition process it is necessary to reflect certain variables, beginning with those whose column totals are initially negative. This is done until all totals add positively.

PROCEDURE IN VARIABLE REFLECTION

There is not just *one* pattern of reflecting that will accomplish this total positivizing of the matrix. Slightly different selections of variables will work and they are equally satisfactory except that some will bring out a little more of the total variance within the first factor than will others. There is no especial virtue in bringing a great deal out in the very first factor; one has to proceed through as many extractions as there are factors *however* they may be brought out. These *initial* reflections can be recorded by actually changing the names of the variables reflected, e.g., writing "unsociable" for "sociable," and keeping to these throughout. Alternatively, it is more systematic to retain the labels and record the sign changes in the factor matrix, on which also the changes later made on other factors, by the method shortly to be described, are recorded. The important thing is that all such sign changes be carefully *recorded*, for experience shows that failure to do so is a common source of error.

Let us suppose that the experimenter has provided himself with an adequate supply of correlation matrix sheets of the design shown in Table 1, p. 41, and a factor matrix sheet as in Table 8, p. 62 (with an extra set of columns for recording signs) and that he has copied out the obtained r's in duplicate in the former, above and below the blank diagonal line (in order that he will not have awkwardly to turn a corner in adding all the r's for one variable). This matrix he will

label R_o, the original correlation matrix, so that the subsequent R numbers will correspond to the number of factors extracted. Beginning with this matrix as given by experiment, he will now reflect those variables with a negative total, as stated above, and in this first matrix he may make the sign changes on the actual matrix by erasure if they are few, or by the method described below for all subsequent reflections if they are many. Naturally he will be careful to change the sign back where two reflections intersect and to reflect for each variable both its row *and* column, as illustrated in Table 14.

TABLE 14.

Given correlation matrix, R_o

	1	2	3	4	5	6
1. Reaction time	O	−.4	.1	−.2	.4	.2
2. Rigidity	−.4	O	−.1	−.1	.5	−.2
3. Intelligence	.1	−.1	O	.2	.1	.6
4. Sociability	−.2	−.1	.2	O	−.4	.4
5. Suggestibility	.4	.5	.1	−.4	O	.1
6. Reading speed	.2	−.2	.6	.4	.1	O
Totals	.1	−.3	.9	−.1	.7	1.1

Reflected matrix, R_o^1

	1	2(−)	3	4(−)	5	6
1. Reaction time	O	.4	.1	.2	.4	.2
(−)2. Flexibility	.4	O	.1	−.1	−.5	.2
3. Intelligence	.1	.1	O	−.2	.1	.6
(−)4. Unsociability	.2	−.1	−.2	O	.4	−.4
5. Suggestibility	.4	−.5	.1	.4	O	.1
6. Reading speed	.2	.2	.6	−.4	.1	O
Totals	1.3	.1	.7	−.1	.5	.7

In the illustration variables 2 and 4 had negative totals so both were reflected. (Note that −0.1 at their intersection does not change its sign nor would the communalities if written in the circles at the intersection of the row and column for each.) As it happens, the simultaneous reflection of 2 has made the reflection of 4 turn out to

have a negative total, so 4 would be reflected back again, giving a total for column 2 of +.3 and column 4 of +.1 and the rest all +ive. This mutual influence of reflections needs constant watching.

ESTIMATION OF COMMUNALITIES

Now comes the step of choosing communalities to fill in the blank diagonal, a rough method for which (No. 1 below) has already been given in Chapter 3. No royal road to choosing the ideal communalities exists and some theoretical digression is necessary at this point to give the pros and cons of the various methods proposed. The chief of these latter are:

1. *Method of highest correlation.* Here one enters (with positive sign) the largest r (positive or negative) that the test has with any other test, i.e., the largest r in the column. Thus in Table 14 a value of 0.4 would be inserted in the circle at the top of column 1. This is the simplest rule-of-thumb device and is widely used.

The reader will remember that the communality is the correlation of a test with itself due to the *common factors alone.* It is written h^2 and equals $1\text{-}S_u^2$ where S_u is the loading in the specific or unique factor (plus error), i.e., it is complementary to the amount of the specific. Now the correlation of two *different* tests together is equal (see page 51) to the products of their separate correlations with that general factor which they both share. This is equal to the products of the square roots of their communalities if only a single common factor is involved, because $h_a^2 = r_{ag}^2$ and $r_{ab} = (r_{ag})(r_{bg})$. When the correlation of the two tests is due to their having several factors in common, the relation is not so simple. If the factors are a, b, c, and d, the communality for test j, namely h_j^2, now equals $a_j^2 + b_j^2 + c_j^2 + d_j^2$, and the correlation of j with k equals $a_j a_k + b_j b_k + c_j c_k + d_j d_k$. In the first case if the two tests are about equal in their common factor loading, i.e., $r_{ag} = r_{bg}$, their intercorrelation will be the same as h^2 for either. In the second case all four of the factor loadings will have to be about the same for each (a less likely situation) for h_j^2 to equal h_k^2 to equal r_{jk}.

Nevertheless it remains true that the correlation of any two tests is about equal to the communality of either or to their mean communality. However, to say that a test with a highest r of 0.8 *must* have 0.8 communality because there is *at least* one other test which

shares this with it cannot be quite correct, for generally the two tests cannot be assumed to have equal communality. There is always one which stoops down and one which reaches up. The r between them is about halfway between their communalities, and we have no means of finding out definitely which is the lower.

However, it is possible to get some idea as to which is contributing more by observing which test is high in its *other* correlations too. For a single high r in a column of lows is not so convincing an argument for high communality as a high r topping a series of high r's. If we arrange the variables in the order of their mean r's with their fellows, therefore, we should expect that the variables of middling mean r's will have r's about equal to their communalities; but the highest r's found for those with high means will underestimate their communalities, and conversely for the highest r's of variables of low mean r. (Incidentally this is similar to Galton's laws of filial regression: that tall fathers will have sons shorter than themselves and short fathers will have sons taller than themselves, because the chances are that a man of very deviant stature will not mate with a wife quite as deviant as himself.)

2. *Method of modified highest correlation.* The considerations just discussed lead to the practice suggested by Burt (**11**) of estimating the communalities of tests with high mean r's to be a little higher than their highest r's and of those of low mean r's to be a little lower than their highest. This parable of the talents treatment, however, must itself be modified a little by recognition that the values of all r's are affected by error. It is generally recognized that in scattered measurements in which some fraction of the variance among them is due to error, the higher values will include more upward than downward error and vice versa. The extreme case is extreme partly because of a genuine extreme score and partly because the chance error is in a direction away from the mean. This is evidenced by the fact that on remeasuring more extreme groups, i.e., redistributing the chance error, there is generally some regression to the mean—some of the extremity is lost. How much we should plan to tone down the above procedure in allowance for this counteracting effect of error will depend on the known probable error of our r's. Unless this error is great, the practice of dispersing the communalities more than the real r's should prevail. For example, if the mean r in a matrix is 0.4, a

variable with a mean r of 0.6 and a highest r of 0.7 might have its communality estimated at 0.8; while one with a mean r of 0.2 and a highest r of 0.3 might have its communality estimated at 0.2. It is usual, however, to make these allowances without taking the trouble to work out the mean r, assuming that the mean r's and the highest r's will place the variables in the same order.

3. *Use of small clusters.* Where one can pick out from the correlation matrix small clusters of three or four fairly highly intercorrelating variables, the communality of any member of such a cluster can be estimated fairly well by Spearman's formula for a single general factor $r_{ag} = \sqrt{r_{ab}\, r_{ac}/r_{bc}}$. Remembering $h^2 = r_{ag}^2$, we take for the communality the expression on the right without square root. This becomes more reliable if four variables can be involved in a cluster when

$$h^2 = \sqrt[3]{\frac{r_{ab}^2\, r_{ac}^2\, r_{ad}^2}{r_{bc}\, r_{bd}\, r_{cd}}} \tag{19}$$

4. *Use of miniature centroid.* Where it is possible to find larger clusters, say, of four or five tests with high intercorrelation including the test in question, we can perform a miniature centroid analysis in this group. Thus we obtain what is really a first estimate of their true communalities, and these we can insert ready for the centroid analysis of the whole matrix. In this small matrix we first insert communalities estimated by rougher, quicker methods, usually (1) or (2) above, so that each column is complete. The estimated communality for a test a then equals

$$h^2 = \frac{\text{Square of sum of } r\text{'s in completed col. for variables in question}}{\text{Sum of } r\text{'s in the completed miniature matrix as a whole}} \tag{11}$$

A particularly thorough examination of the problem of estimating communalities is given by Thurstone (**126**) who discusses twelve possible methods including three of the above explicitly and the fourth by implication. The final choice of method must depend on (a) what degree of accuracy we are aiming at, i.e., for what purpose the factorization is being made, and (b) what opportunities or restrictions are presented by our particular correlation matrix.

Consideration of the first takes us into the whole philosophy of factorization, wherein many writers wrongly assume that the aim is so to estimate the communalities as to make the number of factors

obtained from the matrix a *minimum*. This view is one of mere mathematical economy, and Thurstone (**126**) has rightly urged against it that the scientific aim is rather to estimate communalities which will help to yield the number of factors most clearly indicated by the rest of the correlations. These are the communalities one would reach after repeated iteration, i.e., repeated refactoring of the whole matrix, using at each fresh start the communalities obtained at the end of the last round, until the communalities no longer change significantly.

One meets also the view that communalities should be estimated as low as possible in order to keep the variance due to the common factors at a minimum. This is another distortion of the principle of parsimony. It reduces the common factor variance only to increase the specific (unique) factor variance. And good arguments will be given later (Chapter 19) for the opposing view that the aim of good factor analysis is ultimately to do away with specific factor variance altogether. For specifics are improbable entities and their large numbers reproach us like a multitude of confessions of ignorance.

However, such discussions on whether one should manipulate communalities to maximize or minimize specifics are largely beside the point if we take the position stated above that (a) the aim is to find the particular number and size of factors already implicit in the off-diagonal correlations, reflecting a true structure in nature, and keep in mind also (b) that, as Thurstone points out (**126**), the effects of overestimating and underestimating communalities are not always those one would naively expect. The insertion of the minimum communalities consistent with a real solution does not necessarily give to the matrix what the mathematician calls its minimum rank, i.e., the minimum number of common factors. Although it nevertheless remains roughly true that giving higher initial estimates to the communalities does yield more factors and larger final communalities, the reverse procedure—attempting to reduce common factors by *reducing* communalities—runs into the danger that when reduced below a certain point they produce what the mathematician calls a non-Gramian matrix. That is to say, the combined matrix of r's and estimated communalities loses the properties of a soluble matrix and yields some factor loadings that are imaginary numbers.

In the present review of methods of communality estimation, the discussion has been restricted to four approaches that are entirely

acceptable even though they differ in merit. The reader may correctly have assumed that it goes without saying that certain other rough methods of which he may have heard are rejected outright. However, we should perhaps pause to mention two of them—namely, putting unity in the diagonal and putting the reliability coefficient in the diagonal—to remove any doubts as to the causes of their rejection.

Putting 1.0 for every communality is obviously assuming something which is definitely incorrect, except in extremely rare experimental designs: The tests do *not* have nothing but common factors; they certainly have some specifics. This practice has the effect of making the rank of the matrix equal to the number of tests, i.e., of making as many common factors as there are tests, as in the principal components solution! The use of reliability coefficients for communalities is less absurd, and it has an apparent, superficial logic about it. For if the value in each cell represents the correlation between the tests, the column and row of which meet there, it seems reasonable that the values in the diagonals should be the correlations of tests with themselves. However, when our purpose is to look for common factors we do not want specifics intruding as common factors, and the reliability coefficient indubitably represents the correlation of the test with itself due to its specific factor as well as its common factor. Moreover, the reliability coefficient is likely to vary with the nature of the population, increasing with heterogeneity for example, in a way having no exact relation to the communality of the test with other tests. Apart from such unrelated variations, the reliability coefficients are systematically too high to give the true communality, on account of the above-mentioned inclusion of specific factor (but not error) variance.

The aim of communality estimation which should direct choice among the above four, or further, methods is that of obtaining communalities which best fit in with the correlation matrix as a whole, giving the rank which *the given correlations* (the off-diagonal correlations) most clearly converge upon in their indications. If our method cannot avoid bringing in systematic error, it would better err by overestimation than underestimation. For apart from some dangers of underestimation, indicated above, notably that of producing a non-Gramian matrix, it can be urged that the complexity of the real psychological or social situations with which we deal is generally

greater than we first imagine. Even when the main factors—of ample variance—are recognized, *it is likely that a considerable number of more remote factors have some influence on the variables.* The series of factors does not end sharply (except in certain simple physical examples, such as is illustrated by the three definite factors in Thurstone's box problem [**126**]), but more frequently it fades off into a number of factors of very faint variance beyond those whose variance is so definite as to affect the rank of the matrix. It is better to inflate these negligible factors a little by higher communality estimates, even though they appear a little distorted, than to lose sections of the substantial factors. This conclusion from a psychological standpoint is also shared by Bartlett (**3**) from the standpoint of a mathematician.

CHOICE OF ESTIMATION METHOD

As it happens, in deciding from the condition of our experiment which of the above methods of estimate is most conducive to good results, we are faced with a happy arrangement of alternatives. If the matrix is a large one, the communality estimate plays so small a part in the column total that the error in it will not normally require an iteration of the whole factor extraction—an iteration which would be onerous with so large a matrix. On the other hand if the matrix is small (say, less than twelve variables), so that the error would be more serious, the necessity of repeating the extraction with continually improved communalities (say, two or three times) does not present a terrible amount of work. Because there are generally many good extraneous reasons (pages 331 and 345) for having an appreciable number of variables, the former alternative in design and computation is generally to be preferred.

By using the simple centroid process, as shown in the main example, worked in Chapter 4, one enjoys the advantage of being able to reëstimate the remaining communality after the extraction of each factor; but in one of the two further methods of extraction described in the next chapter—namely the multigroup method—this is not possible. (See, however, subnote 6, on p. 184.) It therefore becomes important in the multigroup method to adopt the most accurate estimation method obtainable; indeed the success or failure of this otherwise excellent method hinges so completely on a good estimate that it should not be employed in conditions where a good estimate is pre-

cluded, or, alternatively, where reiteration of the first factorization is not routinely practiced.[1]

A comparison of some nine methods of estimating communalities made by Medland (page 318, **126**) under Thurstone's direction shows empirically that the best results were obtained by the miniature centroid method (No. 4 above) and fortunately the multigroup method lends itself to this since one has in any case the preliminary task of finding small clusters for the factor extraction itself. A second best method is No. 3 above, especially if one can take two distinct groups, each containing the variable in question, and average their findings.

With a large matrix—say of 40 to 100 variables—method No. 1, or rather its modification in No. 2, will be more appropriate, since the labor of No. 3 and especially of No. 4 may be too great and would also not recommend itself if one is not using an extraction method already requiring the sorting of variables into clusters. A certain art comes by experience in using this method. Where the highest *r*'s for each of the several columns range from 0.3 to 0.7, one should add about 0.1 to the variables at 0.7 and take 0.1 from those at 0.3, leaving those at 0.5 unchanged and the intermediates changed proportionately. If the *r*'s have rather a high likelihood of error through being based on fewer than 50 to 100 cases, one might extend the higher *r*'s less. A practiced judgment in this method will also gain additional guidance, in all factors after the first, by finding a compromise between the value left as a residual from the last communality estimate (used as on page 53, as a checking device) and the new communality estimated by the present methods for the residual matrix as for the first matrix. The residual communality obtained by subtracting the product matrix from the original matrix is certainly less correct than the communality newly calculated from the *r*'s of the present residual matrix, but it is an independent estimate and should be allowed to affect slightly (perhaps weighing 1 to 4) the new estimate.

In the example in Chapter 4, the communalities were estimated by method 1 though they were rounded off to one decimal place to remind the beginning reader that they are rough estimates (such rounding would not normally be done; even a rough estimate gives some idea

[1] The writer has found this method prone to yield final communalities above unity when estimates are carried out with only average conditions or rough methods.

of the magnitude of the second decimal place). Actually with so small a matrix as used in this demonstration one would not use this method at all, but methods 3 or 4. The use of these two methods may be illustrated on the given matrix (page 41) as follows:

TABLE 15.

	1	3	4	7
1				
3	.28			
4	.36	.04		
7	.06	.04	.43	

$$h_1^2 = r_{1g}^2 = \sqrt[3]{\frac{(0.28)^2(0.36)^2(0.06)^2}{(0.04)\ (0.04)\ (0.43)}} = \sqrt[3]{\frac{(0.0784)(0.1296)(0.0036)}{0.000688}}$$

$$= \sqrt[3]{0.0531661} = 0.376$$

Variables 1, 3, 4, and 7 may be taken as an intercorrelating cluster to illustrate method 3. Here we get $h_1^2 = 0.376$; which compares excellently with 0.37, the value obtained when the whole factorization is completed (page 62).

The miniature centroid method might take these same four variables which would be set up as a small matrix as follows:[2]

TABLE 16.

	1	3	4	7
1	.30	.28	.36	.06
3	.28	.30	.04	.04
4	.36	.04	.90	.43
7	.06	.04	.43	.80
Totals	1.00	0.66	1.73	1.33

Adding the columns and applying formula 2 above we obtain

$$h_1^2 = r_{1g}^2 = \frac{(1.00)^2}{0.472} = 0.212$$

[2] The communalities, to a first approximation, that need to be inserted here are obtained from one of the simplest methods, e.g., highest in column and in this case from Table 1.

In this particular case the result is not so near the true value (or rather the first approximation to it obtained at the end of the first factorization) as is that of the above method, but usually this method is somewhat better.

COMPUTATIONAL CHECKS IN THE CENTROID

With communalities duly estimated by one of the above methods, the matrix—whether it be an original or a residual—is ready for the addition of columns and calculation of loadings as described in the regular process in Chapter 4. All that we need now add in our present review of refinements and additions to basic working procedures is a comment on devices for *checking* centroid calculations. For the student will soon discover that an error allowed to pass in factorization can be far more serious than an error of the same magnitude made in most other statistical calculations. If an error is discovered in, say, the second factor extraction at the time the ninth is being taken out, it is no longer possible to dissect out the consequences of that one error. By the third factor the error has diffused its influence over the whole matrix and seven factors have to be thrown away. (This happens because any column error influences the T value which then influences all loadings.) The source of the most prevalent and serious errors is the reflection process.

Let us first describe two essential and simple checks. First we should plan to add the rows as well as the columns, whereupon the sum values inserted at the right edge of the matrix should match the sum values arrived at along the bottom. This also provides a first check on T, by adding the rows. Second, we can obtain a check on the addition of the column totals to T and the taking of its square root by adding up the loadings *finally* obtained, which should then equal $\sqrt{T}$. Incidentally, with most machines the student will find this calculation more convenient if he works out $1/\sqrt{T}=m$ and multiplies each column total by this.

The next step, calculating the product matrix, requires no cautions except that it is desirable in a matrix from which many factors are to be extracted to carry all multiplications and subtractions correct to three decimal places even though the original correlations are considered accurate only to two places. Secondly, since the subtraction of the product matrix to obtain the residual generally involves subtracting negative and positive from positive and negative values (except in the

very first factor subtraction), computing errors in regard to sign are easily made. Consequently it is better to *reverse all the signs* in the product matrix *and add*. This reversal is simply achieved in the calculation of the product matrix itself by giving the multipliers arranged along the top the *opposite* sign to that which they had when they appeared as loadings from the previous process (while retaining the original signs in the loadings along the vertical edge on the left of the product matrix). It is also a good plan to put the *signs* into the body of the product matrix first (they will form a regular plaid pattern in which any errors will catch the eye) before computing the numerical values, since it is not efficient to attend to numerical calculation and sign considerations at the same time.

In our example (page 52) all the first factor loadings are positive and their reversal along the top only will leave the products wholly negative. The addition of these to the first matrix leaves a matrix with r's about equally positive and negative, which will add to zero (apart from rounding errors, as indicated on page 53). This is a first check on the accuracy of the subtraction.

AN IMPROVED REFLECTION PROCESS

A process which, as already stated, presents difficulties and possibilities of many errors is that of reflection, and we must now give attention to possibilities of rendering the process simpler and more efficient. Holzinger and Harman (**71**) describe a simple process that can be used when one is prepared to carry out the sign changes on the matrix itself, which keeps the original signs in the lower left and the changed signs in the upper right (or vice versa) of the matrix. However, as our example above (page 55) shows, the process of getting all column totals positive often involves reflecting the r's of a particular variable back and forth more than once. With a matrix of any size the repeated erasures and insertions over long rows make working on the matrix itself quite impracticable. An entirely different treatment is then much to be recommended, as follows. First we set aside (erase) the residual communalities which are going to be replaced by communalities estimated afresh from the present matrix by the usual methods. With the communalities out we add each column algebraically. Immediately below this row of totals, which may be called s, we write a row $-\frac{1}{2}s$ obtained by halving and changing the sign of the figures immediately above in the s row. (Turn to Table 17, below.)

Now we look along the $-\frac{1}{2}s$ row and find the variable (number at the head of the column) which has the largest *positive* total. This is V_5 in Table 17. (In the matrix itself this would be the variable with the most substantial *negative* r's, i.e., one that we would first choose to invert.) Mark the *row* in the matrix of this variable's r's and add it in to the $-\frac{1}{2}s$ row. At the same time mark, preferably with red pencil, the actual high positive r which originally caused you to choose this variable for reversal and carry a red line down below it as far as any subsequent rows of calculation may go. (This can only be shown here by changing the column to italics, in Table 17.)

TABLE 17.

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8
v_1		.07	.24	.06	−.26	.00	−.16	−.14
v_2	.07		−.04	.17	.15	−.06	−.23	−.19
v_3	.24	−.04		−.06	−.42	.04	−.03	−.03
v_4	.06	.17	−.06		.16	−.06	−.19	−.16
v_5	−.26	.15	−.42	.16		−.08	−.09	−.08
v_6	.00	−.06	.04	−.06	−.08		.06	.06
v_7	−.16	−.23	−.03	−.19	−.09	.06		.31
v_8	−.14	−.19	−.03	−.16	−.08	.06	.31	
Total $= s$	−.19	−.20	−.30	−.08	−.62	−.04	−.33	−.23
$-\frac{1}{2}s$	.095	.10	.15	.04	*.31*	.02	.165	.115
$+V_5$	−.165	*.25*	−.27	.20	*.31*	−.06	.075	.035
$+V_2$	−.035	*.25*	−.31	*.37*	*.46*	−.12	−.155	−.055
$+V_4$	*.025*	*.42*	−.37	*.37*	*.62*	−.18	−.345	−.215
$+V_1 = B$	*.025*	*.49*	−.13	*.26*	*.36*	−.18	−.505	−.255
$-2B$	*−.05*	*−.98*	.26	*−.52*	*−.72*	.36	1.10	.71
Communality	−.30	−.20	.40	−.10	−.50	.05	.30	.30
Total	−.35	−1.18	.66	−.62	−1.22	.41	1.40	1.01

The process is now to be repeated with the new row just obtained. Look for the highest positive total (which will generally no longer be the variable last inverted), note the number of the column in which it falls, and mark the corresponding row with a check mark and add it to the present row. Do not forget to circle the variable total in red and carry a red dotted line below it through subsequent calculations (represented in Table 17 by continuing columns in italics).

Repeat this addition of rows chosen as having the highest positive total until no positive numbers remain *except for the variable columns which have been reversed* and which now have a red dotted line carried down to mark them. At this point, which we will call the B row, create one further row by doubling the number and changing the sign of everything in the B row. These are the true column totals of the variables in view of the fact that some variables have been reflected. If the estimated communalities are added and the experimenter is careful to give the communalities the same sign as these totals, he will have the usual values for obtaining T and the loadings.

It will be observed that this is achieved without any erasures on the matrix and that the matrix is in correct form to have the subsequent factor product matrix subtracted from it. The rationale of the above procedure should be evident after working and examining an example. The process is illustrated here in Table 17 by the first residual matrix used in Chapter 4 where the reflections were originally carried out (compare) by the literal process of trial and error on the matrix.

Although the communalities chosen are the same as those with the same residual in the reflection process on Chapter 4, page 55, the totals are not identical because v_5 was reflected here in place of v_3 there. Different degrees or manners of reflection do not, however, systematically affect the final rotated factor matrix.

The centroid process of extraction thus proceeds through the following cycle, beginning with the given correlation matrix:

1. Reflection of signs for positive totals (p. 55)
2. Calculation and insertion of communalities (p. 158)
3. Addition of columns (p. 55)
4. Check by addition of rows (p. 161)
5. Calculation of $\sqrt{T}$ and the loadings (p. 56)
6. Check loading total against $\sqrt{T}$ (p. 161)
7. Entering loadings with correct signs in factor matrix (p. 62)
8. Calculation of new product matrix (p. 52)
9. Subtraction from original (or addition with changed product signs) (p. 53)
10. Checking residual matrix for correctness of subtraction and so back to the first process of sign reversal (p. 53)

This cycle repeats itself until a residual appears in which there is nothing of statistical significance left. But before we ask by what criterion one can tell whether anything significant is left, i.e., whether extraction is complete, it is desirable to inspect two further methods of extraction which have special advantages.

Questions and Exercises

1. Name six methods of estimating the communalities in a correlation matrix and describe four of them. Which appear to be the simplest in practice? Why may estimates be incorrect when made by the methods of highest correlation? How may such errors be anticipated and perhaps corrected?
2. What are the effects upon the factorization of (a) putting in communalities of unity, (b) systematically overestimating the communalities, and (c) systematically underestimating the communalities?
3. Describe the adjustments on the machine and the detailed steps by which a residual matrix may be computed with the aid of a standard desk calculator without actually computing a separate product matrix first. Examples or exercises in Chapter 4 may be used for illustration.
4. Given the following correlation matrix:

Variable	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9
v_1		−.35	−.14	.58	.18	.32	−.67	−.02	−.27
v_2	−.35		−.25	−.28	−.09	.45	.28	−.27	.24
v_3	−.14	−.25		.30	.08	−.26	−.11	.08	−.51
v_4	.58	−.28	.30		−.10	.19	−.86	−.25	−.78
v_5	.18	−.09	.08	−.10		−.07	.24	.05	.37
v_6	.32	.45	−.26	.19	−.07		−.35	−.22	−.14
v_7	−.67	.28	−.11	−.86	.24	−.35		.16	.76
v_8	−.02	−.27	.08	−.25	.05	−.22	.16		.12
v_9	−.27	.24	−.51	−.78	.37	−.14	.76	.12	

Using the centroid method as described in this chapter, extract a first factor from this matrix. Choose communalities by one of the methods of highest correlation. Compare your solution with that given below.

5. Using tests Nos. 1, 4, 6, and −7 (i.e., No. 7 reflected) estimate the communality of test No. 1 by Spearman's formula and by the miniature centroid methods. Does the result seem reasonable in each case? How does it compare with your previous estimate?
6. Describe one by one the steps in the process of reflecting variables in a residual to get a positive total without actually erasing and altering signs in the matrix.

7. Describe the main checks that can be applied to calculations in the ordinary centroid method.
8. List ten essential steps in a single cycle of factor extraction by the centroid process. From what experience you have already gained, attempt an estimate of what percentages of the time for the whole cycle are involved in each step.

Solution to question 4.

Factor loadings

Factor Variable	F_1	F_2	F_3	F_4	F_5
v_1	.3	−.7	.0	.5	.1
v_2	−.3	.4	.6	.0	.2
v_3	.5	.1	.0	−.3	−.7
v_4	.2	−.8	.3	.0	−.4
v_5	.7	.1	.0	.0	.4
v_6	−.1	.0	.7	.7	.0
v_7	.0	.8	−.2	−.3	.4
v_8	.1	.2	−.5	.2	−.1
v_9	.0	.5	−.2	.0	.8

The communalities used here were, starting with V_1, 0.84, 0.65, 0.84, 0.93, 0.66, 0.99, 0.93, 0.35, 0.93.

CHAPTER 11

The Clustering Methods of Factor Extraction

In this chapter we propose to describe in working detail two or three methods of factor extraction related to or derived from the centroid, which have the advantage of being quicker. They also have disadvantages, so that a choice among the various methods requires an intelligent appraisal of circumstances, notably of computing skills available. These methods can be used for obtaining, after rotation, either orthogonal or oblique factors and can, of course, eventually yield transformations to correspond to bifactor, bipolar factor, or other preferred constellations. However, we shall carry the process through here toward the goal of an oblique simple structure solution.

THE GROUPING METHODS AND THE ROTATION PROBLEM

Thurstone describes four methods in addition to the centroid, but one of these, the diagonal method, is of only historical interest. A general purpose handbook such as this does best, therefore, to confine itself to three: the *group method*, the *grouping method*, and the *multigroup method*. All three of these shorter methods depend upon finding among the variables in the matrix a sufficient number of acceptable correlation clusters. They economize in the factorization, as it were, by finding the directions in which the correlation structure conspicuously protrudes and they take slices of variance from these strategic directions instead of blindly and laboriously carving the *whole mass* as in the centroid. This enables the variance to be cut down effectively without averaging the r's for *all* the variables at once. An economy of computation is thus obtained which runs through every cycle of the extraction (or the "combined cycles" of the multigroup method).

On the other hand the choice of groups demands a certain art in judgment, so that more skilled assistance is required than in the mechanical centroid process. Even when the factorist is skilled, it is inherent in the method that he takes more of a gamble on estimating the communalities and, in general, runs greater risks of distortion from chance error. Further, the checking methods are not so regular as in the centroid, so that he can go further astray before being alerted to his error.

A feature of the group, grouping, and multigroup methods which may be variously evaluated as an advantage or disadvantage is that they are likely to yield an unrotated factor matrix which is already pretty near to the simple structure position of the rotated matrix. This follows from the fact that when we pick out the clusters, the subsequent factorization tends to give high loadings to the particular variables in each cluster, and low or negligible loadings in the given factor to variables not in the cluster, so that the latter already constitute a hyperplane.

The process is indeed very similar to that division of the matrix into submatrix clusters which occurs in the bifactor method, and which according to Burt's claim gives simple structure directly. However, the group and bifactor methods are nevertheless distinct, for the group methods do not first take out a massive general factor, and in consequence, the present methods are more likely than the bifactor method to yield something near true simple structure. For, it will be remembered, the bifactor process starts off with a large general factor before grouping for submatrices or clusters, and this general factor with no zero loadings is far from representing any possible simple structure, whereas the grouping methods yield *all* factors on the same footing, i.e., each having an appreciable hyperplane.

Although the unrotated factors from these methods are likely to be near simple structure, they are unlikely to be *exactly* at the required position. This is obviously true of the grouping method, which yields orthogonal factors, for simple structure is rarely absolutely orthogonal. But with *any* of the methods we must recognize that some factors do not issue in clusters while *some clusters may represent more than one factor*, so that at least a minority of our factors will not issue from the factor extraction correctly rotated. But the nearness to simple structure is real and presents a substantial gain. For,

as the present writer has found on many occasions, simple structure is obtained with decidedly fewer rotations from this position than from that given by the centroid analysis.

If this saving in rotation time is gained, why did we say above that there are disadvantages as well as advantages to the tendency of the cluster methods to yield a solution already near to simple structure? The disadvantage exists only in the exploratory researches in a new field where the factor structure is not yet known. Since the near-simple-structure position is likely nevertheless to be wrong for *some* factors—a pseudosimple structure—it presents a snare which may prevent the investigator from moving on to the real simple structure. For the first steps toward the true simple structure will in these circumstances generally produce deterioration of the hyperplanes, and unless the investigator realizes that things have to get worse before they can get better, he may never make those bold moves which will locate that *mieux* which is *l'ennemi du bien.*

Simple structure is relative, and in a new field of research it is necessary to explore the degree of definiteness of many possible hyperplanes in wide sweeps before deciding what is good. Consequently it is scientifically preferable when doing basic research in a new area to use the centroid method or else to spin the unrotated matrix from a cluster method at random away from its first position before beginning the search for simple structure. Sets of random λ matrix multipliers, designed to disperse the heavy loadings of the first few factors among all factors, have been provided elsewhere by Landahl (**82a**) and should in general be used with the multigroup and other group extraction methods. Alternatively the trial vector method (p. 204 below) of starting rotation may be sufficiently unbiased for many purposes.

On the other hand, when settling the axes in a realm long structured and known, as in many applied researches, the use of these cluster or submatrix methods (or, alternatively, the employment of a single, direct rotation from a centroid analysis to a known position near the true one) will save much time. The initial prejudice suffered through starting rotation from the special cluster positions, as mentioned above, needs watching principally for the situation where a cluster is specifically due to the *overlap* of two or more factors, producing cumulative effects on correlation, and in this case the cluster will definitely lead one astray from the factor positions.

CLUSTER SEARCH METHODS: RAMIFYING LINKAGE

Since the group, the grouping, and the multigroup methods all require that one first pick out correlation clusters, a word must be said about cluster search methods. In a small matrix of ten to twenty variables, simple inspection suffices to let the clusters spring to the eye; but in the matrices of forty to eighty variables, the mere search for clusters can itself take considerable time unless one has a system. Indeed, as pointed out briefly in Chapter 2, the practice among some psychologists, of seeking to establish unitary traits in the form of clusters instead of factors under the impression that the former are easier to find, fails even in the argument based on laziness or economy! For, with many correlations, clusters are difficult to separate by any sharp criterion from other straggling clusters. Both of these difficulties dog, but need not bewilder, the footsteps of those seeking clusters in large matrices; for in any matrix of a size one would dare undertake to factorize, the cluster search task cannot be as overwhelming as it may be in ordinary cluster search projects which deal with really large numbers of variables taken at random.

It may perhaps hardly be necessary to point out that though some members of a cluster can have negative correlations with other members, one must test whether these correlations are *consistent* by reflecting certain variables to see whether all correlations *could* be made positive. For example, in Table 18, page 171, the following variables—2, 5, 8, 9, 11, and 12—have r's of 0.6 or higher for every possible interrelation among them. If we take a criterion of 0.6 or higher as a condition of entry to a cluster, these six variables constitute a cluster, providing the signs are right. Variables 5 and 11 have negative r's with 2, 8, and 12 which, however, are consistent in that by reflecting 5 and 11 we can make *all* r's positive. But variable 9 has to be omitted in spite of substantial r's, for when we reflect it to become positive with 8 and 12, it becomes negative with 2, 5, and 11, and vice versa.

The simplest method of systematically picking out clusters consists in what has been called the ramifying linkage method (**16**). One adopts a minimum r for right of entry to a cluster—say such as will permit about half the variables in the matrix to be in one cluster or another—and calls any r of this value or higher a *linkage*. It is useful then to circle with a red pencil in the matrix all r's which can be

considered linkages. Now one begins with the first variable and looks along the row, writing down the numbers of the variables which have linkages. In the example which follows (shown in Table 18), adopting an r of 0.5 or higher as a linkage, we should find only one for the first variable and so should turn to the second and write down

2 with 4, 5, 7, 8, 9, 11, 12.

We should then take 4 (the first of the set on the right) and try it with the variables to the right of it in this list only. It fails to link with any and so is dropped. We then take 5 and find

5 with 8, 9, 11, 12.

Seven is therefore dropped and crossed out from the initial set, and we try 8, finding

8 with 9, 11, 12.

Continuing in this systematic fashion, we find a cluster of 2, 5, 8, 9, 11, and 12, from which 9 has to be dropped, as shown, for inconsistency of sign.

TABLE 18. Matrix with Cluster Picked Out

Variables	1	2	3	4	5	6	7	8	9	10	11	12
1		.1	−.3	.1	.4	−.3	.5	.4	.2	.1	−.3	−.4
2	.1		.2	.5	−.8	.4	.5	.6	.8	−.4	−.7	.8
3	−.3	.2		.6	−.1	.9	−.4	.6	.5	.4	−.1	.2
4	.1	.5	.6		−.3	.7	.3	.5	−.4	−.5	.3	.2
(−) 5	.4	*−.8*	−.1	−.3		.5	.1	−.7	.9	.2	.9	−.6
6	−.3	.4	.9	.7	.5		.4	−.1	.2	.1	−.4	.3
7	.5	.5	−.4	.3	.1	.4		−.5	−.1	.1	.4	.2
8	.4	*.6*	.6	.5	*−.7*	−.1	−.5		−.8	.2	−.8	.7
9	.2	.8	.5	−.4	.9	.2	−.1	−.8		−.1	.7	.6
10	.1	−.4	.4	−.5	.2	.1	.1	.2	−.1		.4	.2
(−)11	−.3	*−.7*	−.1	.3	*.9*	−.4	.4	*−.8*	.7	.4		−.7
12	−.4	*.8*	.2	.2	*−.6*	.3	.2	*.7*	.6	.2	*−.7*	

Note: Finally accepted "linkages" shown in italics in lower half.

Now that the means of picking out clusters have been illustrated we can proceed to show how the obtained clusters are used in the various shorter factor extraction methods beginning with the simplest —the grouping method.

GROUPING METHOD

Mark out with red pencil frames (within the matrix frames), the rows and columns for the variables in the cluster as shown in Table 19 (A). (In printing here the rows and columns that would be drawn in red frames are shown in italics and with an asterisk at the end of each.) Incidentally the cluster should not have too few variables (say not less than four) nor be too large relative to the matrix. Six to eight constitutes a good number from which reliable communalities can be estimated, but in a matrix of 80 to 100 variables one might take as many as 12 or 14 into a cluster. The size of the clusters can be controlled by accepting higher or lower limiting r's for admission to the cluster in question.

Change right along the row the signs of those variables (5 and 11)

TABLE 19. Extraction by Grouping Method

A. Cluster made consistent by having matrix with 5 and 11 reflected, showing addition process

Test	1	2*	3	4	−5*	6	7	8*	9	10	−11*	12*
1		.1	−.3	.1	−.4	−.3	.5	.4	.2	.1	.3	−.4
*2	*.1*	*(.8)*	*.2*	*.5*	*.8*	*.4*	*.5*	*.6*	*.8*	*−.4*	*.7*	*.8*
3	−.3	.2		.6	.1	.9	−.4	.6	.5	.4	.1	.2
4	.1	.5	.6		.3	.7	.3	.5	−.4	−.5	−.3	.2
*−5	*−.4*	*.8*	*.1*	*.3*	*(.9)*	*−.5*	*−.1*	*.7*	*−.9*	*−.2*	*.9*	*.6*
6	−.3	.4	.9	.7	−.5		.4	−.1	.2	.1	.4	.3
7	.5	.5	−.4	.3	−.1	.4		−.5	−.1	.1	−.4	.2
*8	*.4*	*.6*	*.6*	*.5*	*.7*	*−.1*	*−.5*	*(.8)*	*.8*	*.2*	*.8*	*.7*
9	.2	.8	.5	−.4	−.9	.2	−.1	.8		−.1	−.7	.6
10	.1	−.4	.4	−.5	−.2	.1	.1	.2	−.1		−.4	.2
*−11	*.3*	*.7*	*.1*	*−.3*	*.9*	*.4*	*−.4*	*.8*	*−.7*	*−.4*	*(.9)*	*.7*
*12	*−.4*	*.8*	*.2*	*.2*	*.6*	*.3*	*.2*	*.7*	*.6*	*.2*	*.7*	*(.8)*
Σr for Cluster	0.0	3.7	1.2	1.2	3.9	0.5	−0.3	3.6	0.6	−0.6	4.0	3.6
First Factor Loadings	.00	.78	.25	.25	.82	.10	−.06	.76	.13	−.13	.84	.76

$$T=23.2;\ \sqrt{T}=4.72;\ m=\frac{1}{4.45}=0.21$$

Note: As indicated above, the italics in Table 19 show rows used in addition. The asterisks mark the corresponding tests.

B. First residual matrix

Test	1	2	3*	4*	5	6*	7*	−8*	9	10	11	12
1		1.	−.3	.1	.4	−.3	.5	−.4	.2	.1	−.3	−.4
2	.1		.0	.3	−.2	.3	.5	.0	.7	−.3	.0	.2
*3	−.3	.0	(.9)	.5	.1	.9	−.4	.8	.5	.4	.1	.0
*4	.1	.3	.5	(.7)	−.1	.7	.3	−.3	−.4	−.5	.5	.0
5	.4	−.2	.1	−.1		.6	.1	.1	1.0	.2	.2	.0
*6	−.3	.3	.9	.7	.6	(.9)	.4	.2	.2	.1	−.3	.2
*7	.5	.5	−.4	.3	.1	.4	(.5)	.5	−.1	.1	.3	.2
*−8	−.4	.0	.8	−.3	.1	.2	.5	(.8)	−.7	−.3	.2	−.1
9	.2	.7	.5	−.4	1.0	.2	−.1	−.7		−.1	.8	.5
10	.1	−.3	.4	−.5	.1	.1	.1	−.3	−.1		.3	.3
11	−.3	.0	.1	.5	.2	−.3	.3	.2	.8	.3		−.1
12	−.4	.2	.0	.0	.0	.2	.2	−.1	.5	.3	−.1	
Σr for Cluster	−.4	1.1	2.7	1.9	.8	3.1	1.3	2.0	−.5	−.2	8	.3
Second Factor Loadings	−.10	.29	.70	.49	.21	.81	.34	.52	−.13	−.05	.21	.08

$$T=15.1;\ \sqrt{T}=3.89;\ m=.26$$

C. Factor matrix

Test	F_1	F_2
1	.00	−.10
2	.78	.29
3	.25	.70
4	.25	.49
5	(−).82	(−).21
6	.10	.81
7	−.06	.34
8	.76	(−).52
9	.13	−.13
10	−.13	−.05
11	(−).84	(−).21
12	.76	.08

Note: The loadings with sign reversed because of reversal of the variable have signs in parenthesis.

which need to be reflected to be positive with the majority in the cluster. Make these sign changes in anticipation on the factor matrix, Table 19 (C) (in which there are not yet any numbers). As in the centroid method, these changes will affect the signs of this factor loading and of all subsequent factors as far as this variable is concerned—unless the signs in the matrix are reflected back before the end of the extraction.

Estimate the communalities for the members of the cluster only, preferably using the miniature centroid method (page 160), and insert these in the diagonals.

Add up, as in the centroid method, the r's in the columns. This is done for every column (every variable) in the matrix. But it operates *only upon the rows of the actual cluster members*, i.e., those marked in the horizontal frames: 2, 5, 8, 11, and 12. (There will be no communalities in the other rows.)

The column totals are then treated exactly as in the centroid method, adding the whole row with neglect of signs (some will be negative), and multiplying through by $1/\sqrt{T}$ as in our first example as shown in Table 19.

A product matrix is then calculated using the true, final signs of the loadings (but reversed all along the top edge as usual, if addition is to replace subtraction in the next step) and this is subtracted from the matrix, member by member, as usual.

The residual, as shown in Table 19 (B), must then be inspected to find another good cluster. Such a cluster is at once evident in 3, 4, and 6 which were already prominent in the original matrix. To these, in order to make a more substantial cluster, 7, and 8 (reversed) are also added. The reflection and addition process for the second factor is shown in Table 19 (B) along with the resulting factor matrix for the first two factors (C).

GROUP METHOD

This is, in practice, if not at all points in theory, very similar to the grouping method from which it differs principally in (1) using a double step in finding the cluster, the first being only a tentative searching device, and (2) putting more emphasis on the cluster in that *only* the variables in the cluster receive any substantial loading in the factor being extracted. Some regard the first as contained also in the definition of the grouping method.

TABLE 20. Illustration of Group Method

Tests, # (showing cluster with three tests reflected)	1	2	3	4	5	6	7	8	9	Σ rows	$s\times\Sigma$ rows	$u\times\Sigma$ rows
(−) 1	(.50)	−.35	.02	−.46	.09	.49	−.08	.63	−.21	0.63	−0.63	−0.63
2	−.35	(.25)	.00	.30	.00	−.35	.00	−.45	.15	−.045	0	−0.45
3	.02	.00	(.05)	−.08	.18	.00	−.16	.00	.00	0.01	0	0
(+) 4	−.46	.30	−.08	(.60)	−.36	−.42	.32	−.54	.18	−0.46	−0.46	−0.46
5	.09	.00	.18	−.36	(.55)	.00	−.72	.00	.00	−0.26	0	0
(−) 6	.49	−.35	.00	−.42	.00	(.50)	.00	.63	−.21	0.64	−0.64	−0.64
7	−.08	.00	−.16	.32	−.72	.00	(.35)	.00	.00	−0.29	0	0
(−) 8	.63	−.45	.00	−.54	.00	.63	.00	(.45)	−.27	0.45	−0.45	−0.45
9	−.21	.15	.00	.18	.00	−.21	.00	−.27	(.20)	−0.16	0	−0.16
Σr	2.83	1.85	0.49	3.26	1.90	2.60	1.63	2.97	1.22			
s (weights)	−1	0	0	+1	0	−1	0	−1	0			
$\Sigma(r)(s)$	−2.08	1.45	−0.10	2.02	−0.45	−2.04	0.40	−2.25	0.87	−2.18	−2.18	
u	−1	1	0	1	0	−1	0	−1	1			
$\Sigma(r)(u)$	−2.64	1.85	−0.10	2.50	−0.45	−2.60	0.40	−2.97	1.22	−2.79	⟷	−2.79
$(m)(\Sigma_{r)(u)}=a_{ij}$	−.71	.50	−.03	.67	−.12	−.70	.11	−.80	.33	−.75		

$$u=\begin{cases} 0 \text{ for } & |\Sigma(r)(s)|<\dfrac{2.25}{3}=0.75 \\ +1 \text{ for } & \Sigma(r)(s)>0.75 \\ -1 \text{ for } & -\Sigma(r)(s)>0.75 \end{cases}$$

$-0.75=\Sigma(u\times\Sigma a_{ij})$
= expected sum of weights

$$T=\Sigma u(\Sigma(r)(u))=(-1)(-2.64)+(1)(1.85)+(0)(-0.10)+1(2.50)+(0)(0.45)+(-1)(-2.60)+(0)(0.40)$$

$$+(-1)(-2.97)+(1)(1.22)=13.78 \qquad \sqrt{T}=3.7; \qquad m=\frac{1}{\sqrt{T}}=0.27$$

Test	F_1	F_2
1	.1	.7
2	0	−.5
3	.2	0
4	−.4	−.6
5	.9	0
6	0	.7
7	−.8	0
8	0	.9
9	0	−.3

These changes involve the following detailed differences in procedure. A *small* cluster of three or four variables is first chosen either by inspection as in the previous method or by summing the absolute r's for each column, taking the variables with the highest total and picking out the two or three variables which have the highest (mutually consistent) correlations with this pivot variable. One then reverses signs for the variables, if any, that are negative in the cluster, marks out the rows for the cluster variables, and adds, as in the grouping method (after estimating communalities). *Any other variables that now prove to have an appreciable prospective loading in this factor*, as shown by a column total not less than, say, a third of the general level for the original variables in the cluster, *are added to the cluster*. Thus in the example below, variable 4 is the pivot test and variables 1, 6, and 8 are added to it to form the initial tentative cluster, the last having their signs reversed. Now, from inspection of the totals, as shown at $\Sigma(r)(s)$, i.e., the sum of the r's for the "special" rows in the preliminary pilot cluster, variables 2 and 9 would also be added. Their introduction to the cluster is shown by the new weightings at u, i.e., the "ultimate" cluster, and the subsequent loadings a_{ij} derived from $\Sigma(r)(u)$. In Table 20, as in Table 19, the members of the cluster, in terms of rows to be added, are shown by italicized figures, and would be shown in the actual working matrices by red lines "framing" the rows to be added and guiding the eye.

Below and at the right of the heavy ruled lines we have the check for our example. Looking at the process in more detail we see that in the first line below the body of the table—Σr—we have found the sums of the numbers in each column without regard to sign (or the sums of their absolute values). The next row, marked s and representing the

weightings, is found by entering 1 for the columns of those variables, the rows of which show that they are in the chosen cluster and which are not reflected. A -1 is put in the row for those that are reflected and 0 for those not in the cluster. These numbers are then used to multiply the corresponding r's in each column before adding them. For this purpose we can put the signs to the left of the rows concerned: 1, 4, 6, and 8. Thus in column 1 we have:

$$\begin{aligned}&-1(0.50)+0(-0.35)+0(0.02)+1(-0.46)+0(0.09)\\&-1(0.49)+0(-0.08)-1(0.63)+0(-0.21)=-2.08\end{aligned} \tag{20}$$

As indicated above, the actual addition is best carried out by drawing frames around the rows with a 1 loading, that the eye may neglect the 0 loaded rows, and by reversing the row of signs in the matrix for the tests with a negative sign weighting.

These totals are recorded in row 3 marked $\Sigma(r\times s)$. Now, we set up columns to the right of the body of the table for checking. In the first of these we find the sums of the entries in each row taken with their signs. In the next column we have the products of the sums just found with the corresponding s's from the second row at the bottom. The sum of this second column should agree within rounding errors (in our example, the agreement is exact) with the sum of row three of the check. Now we use the arbitrary criterion previously referred to in finding the largest numerical entry in row three (-2.25) and dividing it by 3 to find a point of reference for the weights of each column. The result here is that anything above 0.75 in numerical value is taken for inclusion in the cluster. Accordingly, we now fill in row four of the check designated u, where under each total of the preceding row larger than 0.75, we place 1. Next, for row five, we use row four in the same manner as we used row two to get row three, namely by multiplying each entry in a column of the table by the u whose column number is the same as that of the row of the individual table entry being multiplied. Again we add these results by columns to obtain row five, and the sum of the numbers in row five should check with the sum of the numbers in column three of the check obtained by multiplying each entry of column one of the check by the corresponding u from row four.

To complete the check, we calculate m as indicated below the table. T is the sum of all the products formed by multiplying the entries of row 5 by the corresponding u's. We extract the square root of T,

take its reciprocal, and this is the value of m, which is used to multiply each entry of row five to obtain the corresponding entry of row six. The sum of row six should finally compare closely with the theoretically expected sum of the weights found by adding the columns of the body of the table with regard to signs and multiplying each column total by its corresponding u, adding the results to find the expected sum, multiplied by m.

With this enlarged cluster perhaps covering half the variables, one then proceeds, as indicated above, to insert frames (heavy dotted underlinings) and add, as in the grouping method, after first reflecting signs to make those variables which began by being negative with the group consistently positive with all.[1]

The second difference of the group from the grouping method which now becomes evident is that the absolute sums of the column totals (each of which is commonly referred to by the symbol t) required to get T is taken *for the members of the cluster only*, i.e., for the column totals of Nos. 1, 2, 4, 6, 8, and 9 in the example of Table 19. *But the loadings are obtained by dividing **all** column totals* by $\sqrt{T}$.

After subtraction of the product matrix in the usual way, the above process repeats itself to completion as in the grouping method. The checks, as shown in the example of Table 19, are also parallel to those of the grouping method.

THE MULTIPLE GROUP METHOD

This method differs from any yet met by the student in that it extracts all factors simultaneously instead of successively. There is no cycle of products and residuals. However, the process necessarily has some equivalent whittling down of r's by extraction of the various factors, and this necessity makes it impossible to proceed to completion in a single step. One must first have a guess at the number of factors to be taken out and then take out at one step somewhat fewer than this. If any residual variance persists (and it generally does if one has been careful not to overshoot the mark) some additional factors are taken out. These additional factors are again removed but this removal

[1] This reflection can be achieved by putting signs (weights) of 1 or —1 at the end and foot of the columns, but it is too much for most computers to notice these calls for reversal as they add the columns, and it is preferable to change them on the matrix itself or to use the method described on page 163.

of batches of factors has to proceed tentatively until no residual remains.

The steps for the *multiple group centroid analysis* may be set out systematically as follows:

1. Choose a number (k) of clusters, each as independent of the others as possible, using any convenient method. Normally they will be what we have called phenomenal clusters, picked out for the first time. But occasionally there may have been prior group centroid analysis of the same or a similar matrix; and the same clusters as were then used for the successive groups may be used for the simultaneous groups here.

   If one finds, when the first set of factors corresponding to these clusters have been worked out, that some appreciable residual remains, a guide to what variables are likely to need inclusion in further factors can be gained from inspecting the communalities of variables (h^2) as obtained from squaring and adding their loadings on the existing factors. Other things being equal, those with low h^2 values are likely to make up new clusters.
2. Estimate communalities for each variable which appears in any one of the clusters, using any convenient method. For example, they may be the obtained values from a previous analysis or they may be the highest r in the column of each member of the cluster within the cluster, but preferably they will be obtained by the miniature centroid method, since a good estimate of h^2 values is very important in this method. Enter these estimates along the principal diagonal of the matrix.
3. Using the procedure for the regular *group* centroid method and using each of the k clusters in turn, compute loadings for what would be the first factor by the group method except that it is here done for all the factors. Employ all the usual checks for accuracy, but do not compute either product or residual matrices.
4. Assemble the results of these applications in the form of a factor matrix having k columns and n rows (n being the number of variables). This matrix may be called $V_{o.o}$, and it contains the *oblique, unrotated factor loadings*. Previously, unrotated factor matrices have been orthogonal, but this method of extraction produces them oblique.
5. As we shall see below in the discussion of the meaning and utility of the present method, it *is* possible to stop at process 4 and to

rotate to simple structure *from the oblique* factor matrix (which we have called $V_{o.o}$). But this is not generally recommendable, and the next step is therefore to obtain the usual orthogonal V_o matrix from this $V_{o.o}$. The reader familiar with rotation is accustomed to V_o being postmultiplied by λ to give an oblique matrix V_o. (If not, two particular aspects of the multigroup method cannot be understood until after the next few chapters on rotation have been read.) $V_{o.o}$ is just another oblique matrix, so we need to find the inverse of λ, i.e., λ^{-1} by which to multiply it to get back to V_o. Steps 5, 6, 7, and 8 are directed to finding the required λ matrix. They involve finding first the matrix $\lambda'\lambda$ giving the angles among the factors, extracting λ from this, and calculating its inverse. The steps are as follows:

Assemble the successive values of T from the separate factor column totals of 3 above along the diagonal of a k by k factor covariance matrix. Complete the "off-diagonals" of this matrix as follows. To find the entry for the ith row and jth column refer to the ith successive application, i.e., the ith factor extraction, of the group centroid procedure as in step 3. Using the jth cluster (instead of the ith) to choose the values of t, find the algebraic total of these values of t. (It will be recalled that the t's are sums of individual columns used in computing T.) As a check on this work, compare the two halves of the matrix, which should be precisely symmetric. Call each such entry C_{ij}.

6. Border the factor covariance matrix with the values of $1/\sqrt{T}$ already obtained as part of step 3.
7. Find the factor correlation matrix by multiplying the individual entries of the factor covariance matrix by the multipliers at the head of their row *and* column (two multipliers for each entry, i.e., $(1/\sqrt{T_i})(1/\sqrt{T_j})C_{ij}$). Again check the matrix for symmetry.
8. Apply the diagonal method of factorization (discussed in more detail in Thurstone (**126**)) to the factor correlation matrix as follows. All the communalities are taken as 1.000. To find the loading of the ith factor variable (corresponding to the ith row) on the jth diagonal factor proceed as follows:
   a. The loadings must be found by columns, proceeding from the top to the bottom of each column in turn.
   b. All the loadings *above* the diagonal of the matrix resulting from the diagonal factorization will be zero. (This is an ex-

ample of a triangular matrix such as will be encountered again later.)

c. The first loading to be found in each column will be the diagonal value. To determine it:

   i. Enter 1.0000 in the dials of a calculating machine. (Upper dials on a Friden machine, lower dials on a Monroe. Here follow through the discussion as on a Friden calculating machine.)

   ii. Subtract the square of any loading already assigned to the given factor variable in the diagonal factor matrix using the negative multiplication.

   iii. Extract the square root of the remainder. The positive value of this root is the desired value.

d. Determine the remaining values in this (the jth) column in any order as follows:

   i. Enter the correlation between factor variables i and j given by the factor correlation matrix in the upper dials of the Friden. Enter this figure positively regardless of its sign.

   ii. Set up the keyboard to form the product of the loadings of variable i and j on each preceding diagonal factor in turn. Combining the signs of these loadings with the sign of the original correlation, *use negative multiplication if the number of minuses is even and positive multiplication if the number of minuses is odd.*

   iii. After performing step ii for each of the preceding diagonal factors, a complement figure may stand in the upper dials of the Friden. If such is the case, this complement is converted to true figures. If not, the true figures are allowed to stand.

   iv. The true figures standing in the upper dial are divided by the entry in cell jj of the diagonal factor matrix. The sign of the correlation originally entered is now attached to this quotient unless a complement was converted in step iii. This result is now entered in the ith row and the jth column of the diagonal factor matrix.

9. The next step consists in calculating the inverse—(λ^{-1})—of the triangular factor matrix λ just obtained. The calculation of an inverse for a symmetrical matrix is set out in Chapter 13, page 226, and it is a formidable task. But the present computation presents

the restricted problem of a *triangular matrix* which is much simpler. (Computing the inverse of an 8 by 8 *triangular* matrix may take about an hour.) Although the calculation of the inverse of a triangular matrix is set out in the chapter on computing, it seems desirable for the convenience of those using this as a workbook to have the computation set out also as an integral part of the multiple group extraction method to which it belongs. The accounts are similar in essentials, but the present is an account of a working procedure developed by Saunders in the writer's laboratory, whereas the more completely set out and illustrated account in Chapter 21 is that published by Fruchter (**56**) and may be read by the student to increase his general understanding of the steps here described.

The present procedure aims to seek the elements of the inverse in the proper systematic sequence, each being solved in turn. In obtaining the elements proceed as follows:

a. Write in all the elements *above* the diagonal of the inverse which will be zeros.
b. Calculate the diagonal elements of the inverse which will be the reciprocals of the corresponding diagonal elements of the triangular factor matrix.
c. The elements of each *row* of the inverse may be found independently starting with the diagonal element and working toward the left, as follows:
d. Prepare a strip which may be aligned with the columns of the diagonal factor matrix.
e. To find the jth row of the inverse, start by writing the jth diagonal element on the jth row of this strip.
f. Align the strip with the $(j-1)$th column of the diagonal factor matrix.
g. Multiply the elements written on the strip by the adjacent elements in the diagonal factor matrix, accumulating the products algebraically.
h. Divide the accumulated products by the diagonal element appearing in the same column of the triangular factor matrix, first converting a negative total if necessary. Attach a plus sign to the quotient if total conversion was necesary, otherwise a minus sign is attached. The result is written opposite the

divisor just used and on the next higher space which is the next lowest available space of the movable strip.

i. The strip is moved one column to the left and steps g and h are repeated.

j. When the left-hand column of the diagonal factor matrix has been used, the top space of the moving strip will be filled. The column of numbers on the strip may now be copied into the *j*th row of the inverse matrix.

10. As a check on steps 8 and 9, the inverse matrix should be multiplied by the factor correlation matrix. The result should equal the diagonal factor matrix within rounding error.
11. The matrix $V_{0.0}$ is now multiplied by the inverse, yielding matrix V_0, which is an orthogonal but unrotated factor matrix. The procedure and checks here are the same as in ordinary rotation.

Notes on the procedure:

1. The method as described does not provide for a significance test of the number of factors.
2. The method as described does not preclude the possibility that imaginary numbers may appear in the triangular factor matrix in one or more columns. If they do, the corresponding rows of the inverse and the corresponding columns of V_0 will contain all imaginary (and pure imaginary) numbers. Rows and columns thus containing terms involving square roots of negative numbers correspond to imaginary factors and will eventually be dropped.
3. If this method is used following preliminary application of some other method, the problem of imaginary numbers (indicating *over*-extraction of factors), as well as the problem of underextraction of factors, should not arise, since one already knows the numbers of factors to extract, and no significance test is necessary for the number of factors.
4. If this method of factorization is the first to be applied, enough factors will have to be extracted, i.e., one must proceed until some *are* imaginary. By temporary permutation of rows and columns in the factor correlation matrix, the introduction of imaginaries into the diagonal factor matrix should be postponed as far as possible. The number of factors obtainable without imaginaries is then the correct number of factors to use.

5. All the work may be carried out on one or two sheets if the calculations are properly arranged.
6. Successive iterations with improved communality estimates may be carried out with a minimum of additional writing out, as a large proportion of the earlier sums found will be unchanged, being independent of the communality estimates.
7. It will be realized by those who have followed the essential steps of the multigroup method that its chief drawback is the special problem it presents at the beginning of rotation. Normally we begin with an orthogonal, unrotated factor matrix, but in this case the factors, as they come out, are already oblique. It *is* possible, as stated, to carry out the successive tentative rotations from the oblique position (found at step 4). The reason that is not recommended is that the results do not represent projections on a *normalized* reference vector unless one goes to the extra labor of multiplying by a triangular factor matrix. (However, for the purpose of obtaining simple structure it is not *particularly* important to have all projections on normalized vectors.) Consequently some may prefer to begin rotations from the oblique $V_{0.0}$ matrix; and later, after the inverse and V_0 have been found, obtain the multipliers for shifting from V_0 to the final V_n. This involves no increase in the number of shifts required and possibly a decrease if the oblique factors are nearer the answer than V_0.

   However, this definitely requires more computing skill, alertness, and experience than does the procedure of putting the $V_{0.0}$ factor extraction result immediately into a V_0 orthogonal matrix and working therefrom in the usual way.

It now remains, after the description of these shorter methods of factor extraction, to discuss ways of choosing intelligently among them in response to the needs and circumstances of various research designs.

Between the group and the grouping method there is little to choose. The former is likely to finish up somewhat nearer to the simple structure position, and it removes the variance in larger slices. If we had really accurate means of testing when a residual is negligible, or of pretesting for the number of factors, the latter would constitute no advantage. For the same number of cycles would be necessary to extract the factors inherent in the matrix whichever method we used. The group method would merely resemble the principal-axis method in

taking out more variance with the early factors and less with the later ones. But since it is not generally agreed by statisticians that a truly effective criterion of the end of the process exists, one is likely to take out more factors by the grouping method than by the group method. The rotation process (Chapter 5) removes these extra factors of the grouping method, but the labor of getting them is lost. Since some people claim that the group method has a more clearly discernible end point, it may enjoy a slight advantage here, though it is always safer to take out too many factors and remove the unreal ones by rotation than to take out too few.

On the other hand, the group method is a little less attractive in its checks and in the economics of computing. It involves two steps in obtaining the cluster, one of which requires the addition of all items in the column, thus putting the process back in the same category as the simple centroid, and one loses the check of adding all loadings to equal $\sqrt{T}$. Choosing the cluster, it is true, is rendered more mechanical, but the choice of clusters in the grouping method presents no difficulty to a normally intelligent person with experience as a computing clerk. The present writer finds the *grouping method* somewhat preferable, especially in the early stages of research in any field and where clusters are not so clear cut in the first correlation matrix that one feels justified in making so sharp a distinction between those variables to be highly and those to be negligibly loaded as is made in the group method. It works best, however, if one selects clusters by the device used in the group method.

The multiple-group methods save computing time over either of the others by omitting the successive residual recordings, though it introduces one counterbalancing extra process. Its chief objections are its complication and its absence of checks. It is probably also more liable to distortion by faulty communality estimates, though this holds to some extent for all cluster methods. On the other hand by this method it is comparatively little trouble to refactor, inserting the communalities obtained from the first factorization; for this does not require one to do the *whole* of the computation again. The attempt to guess the number of factors to be extracted sometimes leads to extraction of too many, and therefore imaginary, factors requiring a forced choice among the possible factors by dropping and rearranging imaginary rows and columns as stated in procedures of step 4 in the procedure notes above. Finally, although the solution is likely

to come out very near simple structure, it emerges, as already emphasized in describing the process, with *oblique factors,* the use of which requires that the computing clerk be able to calculate the inverse at least of a diagonal matrix.

In spite of these objections, multiple factor extraction is a time-saving device in the hands of capable statistical workers and is to be recommended especially where preliminary knowledge of the structure exists. In explorations of entirely new areas—except those of the roughest kind, using the multigroup method with rough communalities—the slightly slower grouping method, or even the simple centroid, leaving the rotation less prejudiced, is probably preferable.

Questions and Exercises

1. State in outline the essential characteristics of the three cluster methods of factor extraction described in this chapter and compare them with the centroid method, particularly in regard to effects upon the subsequent process of rotation.
2. Using the correlation matrix of question 4, Chapter 10, and given the group of tests 1, 4, 6, and −7 (7 reflected) which correlate highly together: find the test among those remaining which, either positively or reflected, correlates best with these. If still another test were to be added to the cluster, which would be the best to choose? Why?
3. Find a cluster in the matrix used in the previous question, which includes tests 2 and 3, either positively or reflected. How can it be easily seen that *both* these tests should not be reflected if used in the same cluster? Which cluster appears to contain more of the variance of the matrix, the one found in question 2 or the one just determined?
4. Pick a cluster of five variables from the matrix of question 2, using a criterion of 0.5 for inclusion in the cluster, and extract the factor which they appear to represent according to the grouping method of Table 15. Note that when two tests in a cluster are reflected, their intercorrelation changes sign twice and hence returns to its original sign in the reflected matrix. Does the factor thus extracted seem to be the same as that found by the centroid method in the previous chapter?
5. Examine the residual matrix obtained in question 4, for another cluster. Does the cluster found in question 3 become more evident at this stage, or less so? How might it be possible for a group of tests showing consistent intercorrelation in one matrix to appear to have lost some of this correlation in the next residual?
6. Using the group method of Table 18, and the same cluster as used in question 4, extract the factor from the matrix and compare the factor loadings and resulting residual matrix with those obtained in question 4. Is there any distinct difference between them? If so, to what may the discrepancies be attributed? Which of the two methods do you consider

to be the better from the standpoint (a) of facility in use (b) of systems of checking, etc.?

7. It is known that the matrix used in these problems has five factors. Following the multiple-group method described in this chapter, find five clusters and extract them simultaneously from the matrix.
8. Discuss the various phases of the multiple-group method in which the statistician's arbitrary decisions influence the results perceptibly. What are the assets of this method as compared with the others studied?

CHAPTER 12

The Elementary Spatial Computations in Rotations

Although in terms of work sequence we should now proceed from extraction methods to tackle the particular problem of how to decide when extraction is complete, this latter must be set aside as a minor matter. For we cannot longer defer carrying the understanding of rotation beyond that elementary stage which the brief exposition of Chapter 5 permitted. Indeed the rotation problem is much more widespread in its implications than is the above relatively restricted problem of finding out when factor extraction is complete. For example, a knowledge of rotation problems and calculations has been required in our discussion in the preceding chapter of the multigroup method of extraction. Indeed, there are many issues in the research design and choice of extraction method, which similarly defy insightful discussion so long as the actual process of rotation remains only vaguely defined in the student's mind. Our purpose, therefore, is to review the meaning and aims of rotation and to explain the algebraic processes by which the actual calculations are carried out. The whole study of rotation developments which will occupy the next four chapters and the present chapter will be concerned simply with the basic calculations, applicable to any kind of rotation.

TEST VECTORS AND REFERENCE VECTORS

It has been shown in the elementary presentation of essentials (Chapters 2 and 5) that the test vectors form a configuration in the common factor space. In this correlation configuration the test vectors remain as rigid in relation to one another as a set of knitting needles sprouting in all directions from an apple. But they are capable of being rotated as a whole with respect to a set of coördinates, and

when some particular approved position has been reached relative to the coördinates of this space we say we have reached a factor resolution (or in Thurstone's terms, a structure). Since the correlation configuration is the real, substantial, *given* thing, and the framework of coördinates is only something imposed from without, it is easier to think (though the whole question is one of relativity) that the former retains its position while the axes are rotated, and this corresponds with the actual working process with the drawings, but one may conceive it either way.

In moving the framework relative to the configuration the student may wonder if he can also shift the origin itself, making what the mathematician calls a movement of translation as well as one of rotation. Strictly he cannot do this. The configuration of vectors has meaning as a set of correlations represented by angles, the cosines of which, taken from the common origin, equal the given r's. Consequently the only movement which retains the meaning of these correlations is one of rotation in which the origin itself remains fixed.

The problem of rotation, as already indicated, is twofold: first, to find a certain rotation position in which the patterns of projections make the best, and usually the only, scientific sense; and secondly to find how to read off accurately the projections on these new axes. A mathematician would suggest that these ends could be attained broadly by one of two methods: (1) geometrical methods, using drawings and models or their equivalents, and (2) algebraic and trigonometrical calculations, by which we could arrive at the new position and projections without actually using graphical devices. It is generally impracticable to get a high degree of accuracy by purely graphical methods, and in the present problem their application to multidimensional rotations is likely to be long and tedious. Consequently, although we kept the initial demonstration in visualizable, graphical form, and although graphical aids are used constantly in most actual working methods, the main procedure is carried along in algebraic expression, which we must now learn.

Graphical aids are least apt in achieving the second of the above two aims of rotation, namely determining the new projections after rotation; but they are almost essential to the first, namely, to *finding* the position which gives simple structure (or whatever particular factor resolution one wants). It is quite conceivable that one could achieve this position by pure calculation, without graphs, e.g., by

setting up some mathematical expression which indicates when the desired position is reached, by attaining a maximum value or surpassing some index figure. There have been attempts at such *analytical* procedures, which will be described later, wherein by the solution of equations one tries to find directly and at once the position to which it is necessary to rotate. But so far the only procedure found to be generally successful by factorists is one which approaches simple structure by repeated trial-and-error gropings. Except in the hands of a visualizing genius, this is best achieved by scanning a succession of actual graphs showing how the test points are moving.

So far we have shown only how the graphs are drawn, looking at the space in two dimensions at a time, and have referred only once to the way in which a change of angle in the drawing can be used to calculate the resulting change of projections. It is necessary now to understand systematically how the calculations of the changes in projection of the test vector end points are related to the changes made in the drawings. For the time being the criterion for shifts on the drawings can be left as already described—a process of trying to find a denser nebula of points through which to run the hyperplane. The latter will appear as an elliptical nebula which gradually hardens into a line as one approaches it in successive drawings.

FIXING VECTORS

To understand how the shift on the drawing leads to calculations in the factor matrix it is necessary to realize how the direction of any vector or axis in space can be fixed by numerical values. First we have to have a given coördinate system with respect to which it can be fixed. In factor analysis this is readily provided by the unrotated factors, which are orthogonal. That is to say, our fixed reference axes are the directions provided by the original centroid axes emerging from the factor extraction process.

Let us consider first the simplest situation of fixing a vector with respect to two coördinates, i.e., assuming it lies in the plane of the paper. To fix a radial line in a single plane we need to know the angles it makes with the two axes which define the plane. The angle with one would alone suffice if it were not that the angle with the second is necessary to indicate on which side of the first the angle is made. Thus after saying that New York has a latitude of 41° from the

equator we have to say what its angle is to the North or South Pole in order that we may know whether it is north or south of the equator. Similarly to fix a vector in *three*-dimensional space we need to know its angles to *three* axes, though the last is fixed in all but sign when we know two. In fact in n-dimensional space we need to know n angles. The student should try this rule by some actual examples in two- and three-dimensional space if it is not convincingly evident.

The angles by which the line or vector is fixed can be recorded numerically in terms of their cosines, which are then arranged in an agreed order in what the mathematician again calls a matrix—a matrix of direction cosines. Thus if F_1, F_2, and F_3 represent the original orthogonal factors, and F'_1, F'_2, and F'_3 represent positions to which we have moved them, then these new positions can be described relative to the old by a matrix of which the following is an example.

TABLE 21. Direction Cosine Matrix

	F'_1	F'_2	F'_3
F_1	−.4	.6	.7
F_2	.4	−.8	−.1
F_3	.8	.0	.7

The reasoning may be followed from Diagram 17, which is an attempt to represent three-dimensional space, and in which only one of the rotated axes, F'_1, is shown in relation to the framework of the three unrotated factors, F_1, F_2, and F_3. It will be seen, by comparison of Diagram 17 and Table 21 that F'_1 has an obtuse angle; α, of cosine numerically equal to λ-0.4 from F_1, an acute angle, β, of cosine 0.4 to F_2, and an acute angle, γ, of cosine 0.8 to F_3. It will be noticed that the sum of the square of the cosines in each column equals 1.00 (as near as can be). This follows from Pythagoras' theorem, as can easily be seen if we take an example in two dimensions only. There, if the rotated axis is of unit length the sum of the squares of the two projections must equal unity, for the axis is the hypotenuse of the right-angle triangle formed by it and the two projections. And by the design of our coördinate systems each axis is always of unit length, representing a total variance of unity in the factor concerned. An exercise in solid geometry will show that the squares of the projections of a unit

vector will also sum to unity in three- or higher-dimensional space. For this reason the projections of each new vector on the old ones, as represented by each column in Table 21, will be found, when squared, to add to unity.

Parenthetically, it may help the student if he notes the difference between what the mathematicians call direction numbers and direction cosines. When the projections are divided by the length of the vector, or when the vector happens to be of unit length, as here, the values placed in the matrix correspond to direction cosines. But if we had vectors of *different* length and simply wrote their projections in the matrix as we do for the tests in the factor matrix, the values correspond to direction numbers, for the latter are projections stated *regardless of the length of the vector.* Naturally one cannot do as many things with the non-comparable direction numbers as with direction cosines, but we should recognize that they exist and that in some situations we use them.

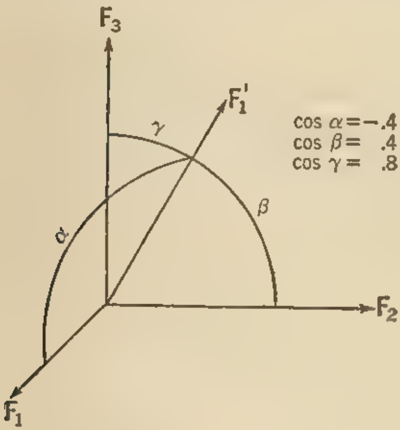


DIAGRAM 17. Position of a Rotated Axis Fixed Relative to Unrotated Axes.

If the projections of a new factor axis upon the old orthogonal ones sum to unity (when squared), it may occur to the student that the projections of one of the *old* axes upon the three *new* ones should also have this property. That is to say, one would expect the values in any one *row*, when squared, to add to unity, just like those in one *column*. This holds if the new axes are all at right angles to one another, i.e., orthogonal, but if each has been rotated independently of the other there is no guarantee that they have finished up orthogonal. They are then said to be oblique axes and the rows will not square to unity.

From the above it will be evident that the directions (positions of the end points) of a set of new axes can always be fixed by a matrix having as many columns as there are axes to be fixed. Each will carry direction numbers fixing the angles to as many orthogonal axes as

exist in the old reference frame. Since the new axes will generally occupy as many dimensions of space as the old axes, the matrix will generally be a square one, with as many cells along one edge as there are dimensions of space. Finally, in such a matrix the columns (and sometimes the rows) will be normalized, i.e., the squares will sum to unity.

CALCULATING ROTATION CHANGES

In rotation the order of procedure is that we first shift the axes—or *reference vectors* as we shall henceforth call them, to indicate their role in relation to the *test vectors*—to new positions. (For the moment we are not concerned with what guides the choice of the given shift.) Having got these new positions we then want to find the projections of the test vectors upon them, from our knowledge of their projections upon the old reference vectors. The positions of the new reference vectors in relation to the old are, as just shown, defined by a square matrix with as many rows and as many columns as there are dimensions of space. This matrix among factors, as illustrated in Table 21, showing the direction cosines of the reference vectors to the orthogonal, unrotated factors, is conventionally referred to as the λ (lambda) or *transformation matrix*. It transforms the old reference vectors to the new ones and, as we shall shortly see, it is also the basis of calculation for shifting the old test projections to projections on the new reference vectors.

Now the calculation of how the projections of a particular test vector change in being transferred from the old axes to the new reference vector will first be illustrated as a process and then explained. First we take that row from V_0, the factor matrix (as obtained from the factor extraction), corresponding to the variable whose new loadings we are going to find. This row is, of course, the specification equation for that variable, and it sets out its loadings or projections upon the existing, unrotated, orthogonal factor axes. For example, let us take a variable v_x which has the following projections on three factors:

$$\begin{array}{cccc} & & V_0 & \\ & F_1 & F_2 & F_3 \\ v_x & +0.3 & -0.1 & +0.6 \end{array}$$

We now want to find its projections on each of three new reference vectors, beginning with F_1', so we take from the square transforma-

tion matrix (Table 21) the column corresponding to F_1' which happens to be:

$$\begin{array}{l} \quad F_1' \\ F_1 -0.4 \\ F_2 +0.4 \\ F_3 +0.8 \end{array}$$

Now we multiply the projection of v_x on F_1 by the projection of F_1' on F_1, i.e., $+0.3$ by -0.4, and similarly the second figure in the row by the second figure in the column and the third in the row by the third in the column, yielding:

$$v_x \text{ on } F_1' = (-0.4)(0.3) + (0.4)(-0.1) + (0.8)(0.6) = +0.32 \quad (21)$$

The new projection is thus compounded of all three of the old projections and this will be seen to be reasonable, since the new axis F_1' has a little of the direction of each of the original unrotated axes, F_1, F_2, and F_3. Moreover, the new loadings partake of the old loadings to the extent that the new axis lies near to the older axis under consideration. For example, F_1' has an r of 0.8 with the old F_3, so that the projection of any variable on the old F_3 should enter a good deal into the present projection. (It does here, to the extent of (0.6) (0.8) the last term above.) But F_1' points *away* from the old F_1 (cosine of -0.4), so any projection the variable had on F_1 should contribute negatively to its projection on F_1'. (It does, to the extent of (0.3) (-0.4), the first term above.)

To what rule will this lead in regard to the whole matrix of factor loadings and the transformation matrix? It will readily be seen that it leads to multiplying every *row* of the unrotated factor matrix (which we have agreed to call V_0) by a *column* of the transformation matrix—in fact by that corresponding to the new reference vector, the projections on which are being obtained. This produces a new column F_1' in the first rotated matrix, which matrix we will call V_1 and which has n variables in it just like the V_0 matrix. The row by column multiplication is repeated for the remaining two columns of the transformation matrix, with the results shown for our standard example (page 195) in Table 22.

If the student happens to be familiar with matrix algebra he will recognize that this procedure follows the regular rules for multiplying one matrix by another. The matrix V_0 is said to be *post*multiplied by

TABLE 22. Multiplying the Unrotated Factor Matrix by the Transformation Matrix

	V_0				λ_1					V_1			
	F_1	F_2	F_3			F_1'	F_2'	F_3'			F_1'	F_2'	F_3'
1	33	−48	21		F_1	1.00	00	00		1	33	−49	−19
2	39	−29	−34	×	F_2	00	71	71	=	2	39	04	−45
3	11	−38	48		F_3	00	−71	71		3	11	−61	07
4	91	−20	−34							4	91	10	−38
5	53	24	−65							5	53	63	−29
6	04	07	16							6	04	−06	16
7	68	47	34							7	68	09	58
8	60	44	31							8	60	09	53

λ_1 to produce V_1. (Postmultiplied because matrix algebra is peculiar in that $a \times b$ is not the same as $b \times a$.) The rule in matrix multiplication is to multiply a *row* of the first matrix by a *column* of the second to produce a column in the product matrix. Multiplying means, as we have seen, that each term in one series is multiplied by the corresponding term in the other and all the results are added algebraically to produce a single figure in the column of the product matrix. Matrix multiplication is possible only when the row of the first matrix has as many values in it as the column of the second matrix. In this case we have n variables or tests and k factors in V_0. Appropriately we have a k by k matrix for the transformation matrix λ. Consequently we finish with a k by n matrix for V_1, the rotated matrix, i.e., each variable is present and each has as many loadings on the new factors as on the old.

PROCEDURE IN WORKING FROM DRAWINGS

The student now realizes how to get the new test loadings from the old, unrotated test loadings, when the axes are spun through any given angles. They are obtained, in the language of algebra, by postmultiplying the unrotated factor matrix V_0 by the transformation matrix λ, consisting of direction cosines corresponding to the angles of shift. But how are these angles to be decided upon from looking at the drawings and how shall we get their cosines for insertion in the λ matrix?

One *can* and sometimes does (as in the example of Chapter 5) draw the graphs from the unrotated factor matrix, make a shift on the

drawing, measure the cosine of the angle, and enter it directly on the λ matrix. (For the moment we shall continue to postpone discussion of just what features in the drawing decide how big the angular movement shall be, though the student knows that in general we are looking for simple structure positions.) Actually the procedure in this shift from V_0 can be treated differently from all subsequent shifts, for it involves writing in only two cosines in the column: that for the axis toward which a move is made and that for the axis from which the new axis is moved. It would be better, therefore, to take a more general case requiring use of the trigonometrical evaluation normally used to give the λ matrix. The reason why this first simple procedure cannot be used later is that as the factors become oblique to one another a move with respect to one becomes a move not merely in the plane of two axes but with respect to all of them. It then becomes necessary to express each shift by calculating some transformation in all the direction cosines of the axis that is shifted, i.e., to change *all* the values in the column of the transformation matrix. Our problem then is to find how a measured angular shift of one axis relative to another, made on the drawing board, will lead to a calculated change in the column of the transformation matrix corresponding to that factor.

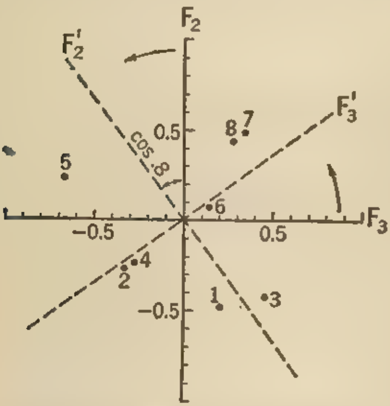


DIAGRAM 18. Projections Before and After a Calculated Shift for Simple Structure.

The rule for this calculation is best first illustrated by an actual example, in Diagram 18, which repeats a shift made in purely graphical terms earlier (page 70) and uses the actual unrotated factor matrix which we extracted by our own calculations (page 62). The test vector end points are first drawn as dots, in the usual way from the V_0 projections, upon the unrotated factor axes drawn vertically and horizontally on the paper. A movement of about 40° on the part of both reference vectors, as shown, will bring more points in the general region of their hyperplanes (considering them both together)

than at present lie there. Actually we shall consider the angle to be 45° in our illustrative calculation, for simplicity of computing. Now to show the general calculations we must set out the transformation matrix values as they exist *before* our shift and *afterwards*, though as this is the first shift the transformation matrix is peculiar. This matrix, λ in Table 23, is really no transformation at all, because the factors are as yet unrotated and lie right along themselves. Thus the direction cosines of F_2 with respect to the orthogonal unrotated matrix are respectively 0 (an angle of 90°) with F_1, 0 (orthogonal) with F_3, and 1.0 with itself (for it lies at 0° to itself). Similarly F_2 and F_3 lie along their own length and at right angles to the remaining factors. Thus it is that the λ matrix of the orthogonal system with respect to itself can be written as in Fig. 1 of Table 23.

TABLE 23. Transformation Matrices

λ

	F_1	F_2	F_3
F_1	1	0	0
F_2	0	1	0
F_3	0	0	1

Fig. 1

λ_1

	F_1'	F_2'	F_3'
F_1	1	0	0
F_2	0	71	71
F_3	0	−71	71

Fig. 2

Incidentally, if the V_0 matrix is postmultiplied by this transformation matrix it will remain exactly as it is, as the student may verify in a moment by application of the matrix multiplication rule—and this is as it should be, since no move has yet been made.

If now we move F_2 through 45°, which, as shown in Diagram 18, will give it a hyperplane nearer to simple structure, the calculation for obtaining the new direction cosines may be written

$$(F_2') = (F_2) - (\tan 45°)(F_3) \qquad (22)$$

where (F_2') represents a function of the whole column of direction cosines for F_2', and similarly for the other factors. The tangent by which the F_3 values are multiplied is *negative* because the movement is *away from* F_3. If the drawing had been one of F_2 and F_1, and if the movement had been toward F_1, the column for the latter would be multiplied by the positive value of the tangent of shift.

It is understood that for the present we are simply learning *what to do* and that when this is familiar the reasons or proof for it can be approached. Let us therefore work out the steps of the calculation thus briefly outlined, still keeping to the example we have carried along from our familiar early calculations.

In an actual computation we should begin by copying out the column of cosines for F_2, the factor on which we have shifted, as shown below. Then, nearby, we should copy out the column for F_3, arranging that horizontally also, for ease of carrying out arithmetical procedures. This F_3 would then be multiplied by $-\tan 45°$, which happens to be exactly -1, and the result would be added to F_2.

$$\begin{array}{rrrr} F_3 = 0 & 0 & 1 \\ (-\tan 45°)(F_3) = 0 & 0 & -1 \end{array} \quad \rightarrow \quad \begin{array}{rrr} F_2 = 0 & 1 & 0 \\ (-\tan 45°)(F_3) = 0 & 0 & -1 \\ \hline F_2 - (\tan 45°)(F_3) = 0 & 1 & -1 \end{array}$$

But these cosines (projections from a hypotenuse of unit length) have to be *normalized*, because we are dealing with a factor of unit length, and the projections of a unit vector when squared must come to unity. When a calculation[1] is carried out to make these values proportionately reduced, so that their squares sum up to unity we obtain the true cosines which are:

$$0 \quad 0.71 \quad -0.71$$

A similar calculation for F_3, where we add (*plus* tan 45°) (F_2) to the F_3 values because F_3 has moved *toward* F_2 (Diagram 18), will yield

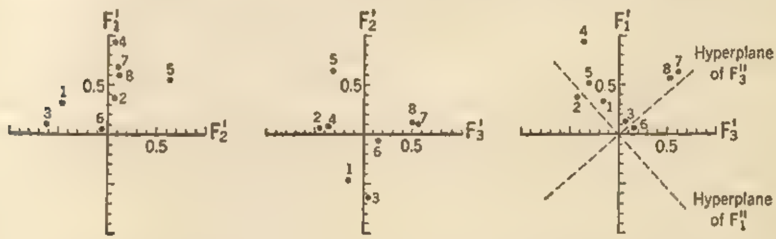
$$0 \quad 0.71 \quad 0.71$$

as the student may quickly verify. The direction cosines for F_1 are not to be changed, because our drawing found no improvement immediately possible in its position. The new transformation matrix, which we may call λ_1, because it will produce the rotated factor matrix V_1, may now be set up by putting these new columns together, as shown in Fig. 2, Table 23; when this matrix postmultiplies the V_0 matrix

[1] The way to normalize a set of numbers is to square them, add these squares, and divide each original number by the square root of this sum. Thus if they are initially a, b, and c, they will become $\frac{a}{\sqrt{a^2+b^2+c^2}}$, $\frac{b}{\sqrt{a^2+b^2+c^2}}$, and $\frac{c}{\sqrt{a^2+b^2+c^2}}$. These if squared and added, give a total exactly equal to 1.

we obtain a V_1 rotated matrix in which the projections on F'_2 and F'_1 are different from the original drawings, in just the same way as was obtained by a purely graphical process before (page 70). This the student may work out as an exercise, comparing the obtained projections of the eight points with those which can be drawn upon the dotted axes in Diagram 18. (The solution for V_1 is given in Table 24.)

The approach to simple structure is by trial and error through several successive steps. Consequently we now plot points in a fresh set of graphs *obtained from the projections given in the V_1 matrix.* We can now see how far the hyperplanes have been improved and we can see by comparing the upright $F'_2F'_3$ drawing (Diagram 19) with $F'_2F'_3$ in Diagram 18 that the calculation has indeed brought the points to the position intended by the shift on the drawing.



Drawing for $F_1'\,F_2'$ Drawing for $F_2'\,F_3'$ Drawing for $F_1'\,F_3'$

DIAGRAM 19. First Drawings Made from Calculated Shift from Unrotated Position.

Now we look for another drawing on which some improvement of hyperplane is possible (obviously $F'_2F'_3$ cannot be improved again until either F'_2 or F'_3 has been improved on another drawing.) Shifting F'_1 toward F'_3 (Diagram 19) looks good and it happens that the angle is again 45°, whence we calculate the new F_1 as follows:

Cosines of F_1' are	1.00	0.	0.
(Cosines of F_3')(tan 45°) give	0.	0.71	0.71
Adding	1.00	0.71	0.71
Normalizing	0.71	0.50	0.50

which are the direction cosines of F''_1. F'_3 is now shifted a similar amount (but *away* from F'_1) since, as the diagram shows, the possibility of a hyperplane with as many as four points in it invites such a shift.

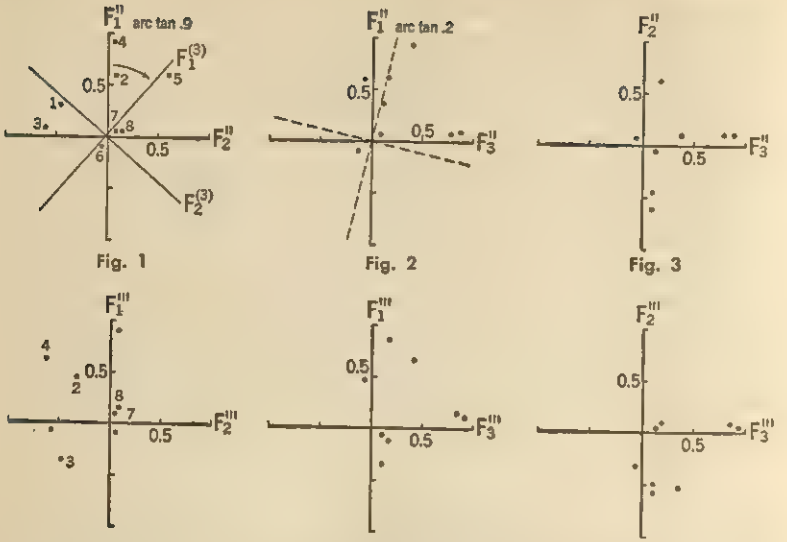


Fig. 1 Fig. 2 Fig. 3

Fig. 4 Fig. 5 Fig. 6

DIAGRAM 20. Second and Third Shifts Drawn from Calculations of New Projections.

The second shift toward simple structure, as shown in the matrix V_2, is obtained by multiplying the *original* V_0 matrix by the new transformation matrix λ_2, which, after the above shifts on F_1 and F_3, is as follows:

TABLE 24. Second Transformation Matrix

λ_2

	F_1''	F_2''	F_3''
F_1	.71	.00	.71
F_2	−.50	.71	.50
F_3	−.50	−.71	.50

The V_2 matrix (see Table 25) now gives the drawings shown in Figures 1, 2, and 3 of Diagram 20. The two transformation matrices and their resultant rotated factor matrices from which the drawings of Diagram 19 and the lower half of Diagram 20 have been made, are set out systematically, together, in Table 25.

TABLE 25.

V_1

F_1'	F_2'	F_3'
.33	−.49	−.19
.39	.04	−.45
.11	−.61	.07
.91	.10	−.38
.55	.63	−.29
.04	−.06	.16
.68	.09	.58
.60	.09	.53

V_2

F_1''	F_2''	F_3''
.37	−.49	.10
.59	.04	−.04
.03	−.61	.13
.92	.10	.38
.58	.63	.17
−.09	−.06	.14
.08	.09	.89
.05	.09	.80

λ_1

	F_1'	F_2'	F_3'
F_1	1.00	.00	.00
F_2	.00	.71	.71
F_3	.00	−.71	.71

A.

λ_2

	F_1''	F_2''	F_3''
F_1	.71	.00	.71
F_2	−.50	.71	.50
F_3	−.50	−.71	.50

B.

A shift on F_1'' and F_2'' as shown in Fig. 1, Diagram 20, indicates itself as a movement still to be made, and a slight improvement on F_1'' in Fig. 2 is also possible, though in these cases it may not be obvious to anyone starting rotation that any substantial improvement to reward one's trouble is obtained. The resulting matrix becomes

TABLE 26.

λ_3

	F_1'''	F_2'''	F_3'''
F_1	.53	−.47	.71
F_2	.10	.86	.50
F_3	−.84	−.19	.50

which when used to postmultiply the original factor matrix gives

TABLE 27.

V_3

	F_1'''	F_2'''	F_3'''	h^2	
1	−05	−60	10	.36	
2	47	−36	−04	.35	Calculation of r_{12} from
3	−37	−50	10	.40	this matrix = 0.19 com-
4	70	−59	38	.98	pares with true value
5	87	07	18	.79	= 0.20, as shown in the
6	−09	01	13	.03	original correlation
7	11	01	89	.80	matrix, page 46.
8	13	03	78	.63	
Number of variables in hyperplane (±.10 or ±.13)	2 or 4	4	3 or 4		

which has as many loadings falling in hyperplanes (within ±0.10 or ±0.13) as it seems possible to get.

Accordingly we may regard this as reaching as good a simple structure as can be obtained, apart from slight reductions which slight shifts will make in the projections of variable 6 on F_3 and 7 and 8 on F_1. For, if we count loadings of 0.10 and less as zero save for chance error, i.e., as lying in the hyperplane, then every one of the three factors has two to four of its variables lying in the hyperplane. If we raise the boundary to ±0.13, since slight polishings of the rotation can bring variables within this boundary into a ±0.10 hyperplane, then every factor has one-half the variables in its hyperplane, which is to be considered, in average circumstances, a very good simple structure.

The student will note that any errors made in the rotation process, e.g., through measuring the shift (in degrees) by a rough instrument or against a too roughly drawn set of points, are *not cumulative*. The calculation returns at every shift to the firm basis of the unrotated factor matrix and comes to a new rotated factor matrix by way of a new transformation matrix.

However, there may be errors in the *last* of these transformations—that leading to the final, simple structure matrix V_n. Against this eventuality we have certain checks which can be applied, at least with an orthogonal rotated matrix. First we may work out the sum

of the squares of the loadings for each variable, which should still equal the same communality, h^2, as in V_0. In this case the calculation shows that these have not altered significantly from the unrotated values, despite all loadings being different, which is as it should be when we rotate but do not move the origin. Secondly, the reader may at this point wonder whether the inaccuracies necessarily existing in getting angles etc. from a graphical method are likely to introduce appreciable errors in the final result. Actually they do not. The position of the points is never so tightly clustered in a hyperplane that

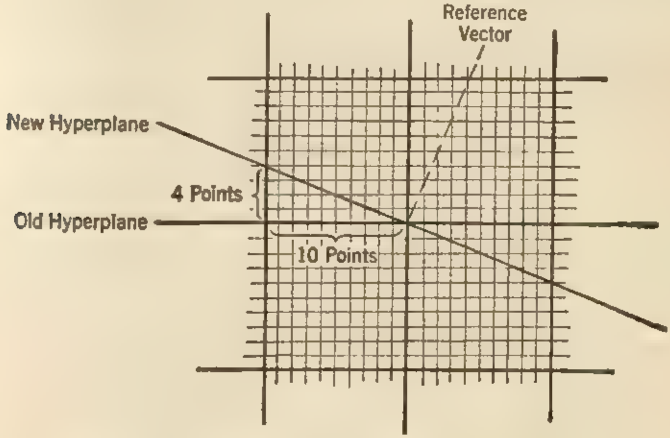


DIAGRAM 21. Reading the Tangent of Shift from the Graph.

an error of two or three degrees in the angle will make much difference in cutting the approximate center of them. The drawings are largely a mere guide to the calculations, which run parallel in a self-contained system. On the other hand, there *are* certain methods of making several graphical shifts in succession before returning to the unrotated matrix, as explained in Chapter 15, which bring cumulative error, but the standard device does not.

In the standard procedure now being described one does not, indeed, usually take a protractor to read off the angle of shift, nor does one look up the exact value of the tangent in a book of trigonometrical tables. Instead it suffices, on all but the very last shift (and sometimes on that too) simply to read off the tangent of the penciled hyperplane on the graph paper itself. To do this one runs along the old hyper-

plane ten points (each usually a 1/10th inch marking) and then counts the number of points (1/10th divisions) upward from that mark until the line drawn for the new hyperplane is encountered, as shown in Diagram 21.

Incidentally, the first one or two rotations as described above are likely to be more difficult than those made when rotation is well started, both in finding the position in the drawing and in making the calculation. The risk of error in the calculations arises from the many zeros in the columns of the transformation matrix and indeed from its generally odd and awkward form (See Table 23). But very soon, as in λ_3 above, there are figures in every row, and the additions of the tangent-multiplied values produce columns in which all positions are regularly occupied by numbers.

TRIAL VECTOR METHOD

However, there is another way of carrying out the first rotation which avoids this cold start, by small shifts, from the unrotated matrix. By this new method one leaps, as it were, immediately to a position remote from the unrotated matrix and presumably nearer to the position one eventually will reach. This is done without a drawing, and the drawings start only when this first trial position has been reached. The calculations are just the same as in the standard method except for the first transformation matrix. This is called the trial vector method.

To calculate the first transformation matrix for trial vectors we begin by deciding on certain positions which we desire the new factors to take up. This is done on extraneous evidence, e.g., our experience in previous factorizations, or a hunch as to which tests are most pure representatives of a factor. The positions of the present test variable vectors are thus employed as a guide. For example, we may expect a general factor of intelligence to emerge in the rotation, and if one of our variables is actually an intelligence test then we may expect that the eventual reference vector for the factor will pass very nearly along this test vector. Or again, we may have a hunch about what particular tests will have *no* loading in one of our factors and consider these test end points to constitute a likely hyperplane, putting our trial vector at right angles to them.

Such insightful and deliberate manipulation is rendered possible because the factor matrix already contains the projections of many

test vectors, and these give us in plenty the direction cosines for diverse possible directions of reference vectors. Thus when we are looking for the column of direction cosines to put in the λ matrix to define the position of the trial reference vector why should we not, after all, use those of some particular test which, as stated in the last paragraph, comes near to the position we believe we want? Incidentally, such use of hunches is not inconsistent with ultimate "blind" rotation.

The direction cosines to be placed in the λ matrix to give the positions of the new trial vectors in relation to the unrotated factors are then to be found in the given unrotated factor matrix, V_0. Actually, the loadings (projections) of a test on the orthogonal axes differ from the direction cosines of a new axis only in that the latter are calculated for a unit length of the axis while the former need to be divided by the length of the test vector (hypotenuse to the projections) which is generally short of unity by an appreciable amount. (The length of the test vector is the communality of the test, which is always short of unity.) If, therefore, we want to put a new axis where one of the test vectors now lies, it is necessary to take the row of loadings for that test from the V_0 matrix and *normalize* it, i.e., make them such that their summed squares are unity, so that the vector is now a unit length axis in the common factor space. These numbers can now be inserted as a column in the λ matrix and used to multiply the whole factor matrix to obtain each test's projections on the new factor. (The particular test used to get the axis will, of course, turn out to have projections *only* on this axis, if axes are orthogonal.)

The details of the procedure can be illustrated by our standard example above. Our objective is to choose from the unrotated matrix (page 62) as many test vectors as there are factors (three in this case) in order to set up axes which at one jump will be decidedly nearer to simple structure than would be our first shifts from the unrotated axes on the old principle of shifting. The criteria that are generally useful in choosing these vectors may be summarized as follows: 1. They should have high communality, i.e., stand out well into the space in which all the tests lie. This is not as essential as the other criteria, for a short test vector *might* lie in the right direction. But in general such a test would not be so useful for describing other tests, and in this case the rule leads us to omit from considera-

tion test 6 and possibly 2 and 3. 2. Looking back next at the original correlation matrix now (page 41), instead of the factor matrix, we should find the test having an adequate number (say one-third) of near zero r's with other tests—for these are tests that would constitute a hyperplane when it becomes a factor. If the test has good communality and yet several near zero r's, it follows that some r's will have to be decidedly high, to make up for the near zeros. In other words, the test should show a *wide range* of r's with other tests. 3. Finally, the three tests we choose must be as nearly as possible uncorrelated among themselves, for they are intended to constitute three independent axes. On these criteria we choose tests 5, 1, and 7 from our example, which are shown in Table 27, first merely picked out and arranged in vertical columns and second as transformed into a true transformation matrix by normalization.

TABLE 28. Transformation Matrix from Trial Vectors.

Selected Test Vectors		t_5	t_1	t_7
Unrotated factors	F_1	.53	.33	.68
	F_2	.24	−.48	.47
	F_3	−.65	.21	.34

λ_1

	F_1'	F_2'	F_3'
F_1	.61	.53	.76
F_2	.27	−.78	.53
F_3	−.75	.34	.38

On multiplying the V_0 matrix by this λ_1 matrix and making drawings, it will be found that the points are in fact so near the final positions by the former process of three shifts that relatively slight shifts are necessary to get simple structure. It is noticed that F_2 in this matrix has opposite signs to those it has in λ_3 (page 201), which means that it is being measured from the opposite end and will load all variables in the opposite sense. This can be rectified, if it makes better sense to measure it from the opposite direction, by reversing all the signs in its column in λ_1 but usually it is a trivial or indifferent matter as to which way the factor is scored.

There is, however, one important respect in which this λ matrix differs from the old one: the factors are *not* absolutely orthogonal. If we wish them orthogonal, as in the former example, we can get them so either by shifts on the graph or by algebraic means. The latter depends on the following concepts. First, we know from the above (page 65)

account of methods of checking a factor matrix against a product matrix that the inner products of the two sets of numbers representing the projections of two vectors upon orthogonal axes yield the correlation (or cosine of angle) between the two vectors. Consequently, if our two vectors (in this case reference axes) are to be at right angles, their inner products must sum to zero. Second, we know that for any variable the sum of the squares of its projection must add to its communality or, in the case of a reference axis, to unity.

In our present example we can take one reference vector with which we are well satisfied, say F_1', and take the *nearest values* to the present F_2' which will give an axis at right angles to it. For example, we could let the first angle in F_2' stand and put b and c for the other two, when we should have two equations:

1. Inner products of F_1' and $F_2' = 0.61 \times 0.53 + [0.27 \times (a)] + (-0.75 \times b) = 0$ (23)
2. Communality of $F_2' = (0.53)^2 + a^2 + b^2 = 1$

The solution of these it will be noticed gives one answer near the present slightly oblique F_2', namely, 0.53, −0.53, and −0.66 and another inapplicable answer. Taking the former as the rectified F_2, the next step is to seek an F_3 at right angles both to the fixed F_1 and the fixed F_2. If we call the required cosines d, e, and f, we have three equations

$$\begin{aligned} 0.61d + 0.27e - 0.75f &= 0 \\ 0.53d - 0.53e - 0.66f &= 0 \\ d^2 + e^2 + f^2 &= 1 \end{aligned} \tag{24}$$

from which to obtain the cosines fixing the required orthogonal axis. But in general we do not go to this further labor of making reference vectors exactly orthogonal, for, as indicated in the next chapter, the oblique factor resolution is likely to be a truer simple structure and hence the preferred interpretation.

The above described leap to a particular set of trial vectors, in the first move of rotation, is only one of several possible preferred starting points, but as Thurstone indicates (**126**) it is probably the most useful and widely used. Alternatives are to pick out two or three variables which one believes should constitute the hyperplane of the factor, or to pick out a cluster of variables and run a reference vector through the middle of them, or to calculate a reference vector at right angles to all of a given set of factor positions. The steps for making these

special moves will be evident from later chapters, though the student may infer them from the last example and the general principles so far studied.

Questions and Exercises

1. What two aspects of a correlation configuration—expressible in two formulas—remain constant no matter what the rotation position?
2. Explain the steps in obtaining the new transformation matrix from the old, beginning with the inspection of the graphic plots, and explain the meaning of the numbers in the matrix. How can one decide from inspection whether a given transformation matrix will or will not yield orthogonal reference vectors?
3. Describe the process of obtaining the V_1 factor matrix from the V_0 matrix and state what calculation checks can be applied when the V_1 matrix is still composed of orthogonal reference vectors. Why are errors in rotation plots or calculations not cumulative?
4. State the rules of matrix multiplication as they are involved in rotation calculations and carry out the following operations with these matrices:

$$M_1\begin{pmatrix} .3 & -.1 & .6 \\ .5 & .2 & -.4 \\ -.8 & .0 & -.3 \\ .2 & -.3 & .0 \end{pmatrix};\ M_2\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix};\ M_3\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix};$$

$$M_4\begin{pmatrix} .2 & .3 & -.7 \\ .0 & .6 & .5 \\ -.4 & -.3 & .0 \end{pmatrix};\ M_5\begin{pmatrix} .1 & -.1 & -.7 \\ .0 & -.2 & .5 \\ -.2 & .1 & .0 \end{pmatrix}$$

   Find the following products by matrix multiplication: (a), (M_1M_2); (b), (M_1M_3); (c), (M_1M_4); (d), (M_1M_5); (e), (M_3M_4); (f), (M_4M_3). Notice that $(M_3M_4)=(M_4M_3)$, and also that the first two columns of (M_1M_4) are multiples of like columns of (M_1M_5). Why?
5. The following factor matrix gives the projections on orthogonal axes at the end of factor extraction from the correlation matrix used in questions in the two preceding chapters.

	F_1	F_2	F_3	F_4	F_5
v_1	.2	−.7	.0	.5	.1
v_2	−.3	.4	.6	.0	.2
v_3	.5	.1	.0	−.3	−.7
v_4	.2	−.8	.3	.0	−.4
v_5	.7	.1	.0	.0	.4
v_6	−.1	.0	.7	.7	.0
v_7	.0	.8	−.2	−.3	.4
v_8	.1	.2	−.5	.2	−.1
v_9	.0	.5	−.2	.0	.8

Calculate the new loadings on F_1 and F_5 if these axes are rotated orthogonally (F_1 beginning as vertical axis) clockwise through an angle whose tangent is 0.67. Does this improve the hyperplanes of the two factors, i.e., do more points seem to fall near the new positions of the hyperplanes than in the unrotated state? Plot the points represented by the loadings in these two factors for each of the nine tests, and draw in the new positions of the axes after rotation. This will serve as a check on the computation as well as to show how the angle of rotation might have been determined, and to show alternative positions of rotation.

6. Plot the nine remaining drawings of the various combinations of factors from the table in question 5, and determine which show a definite suggestion for a rotation. Are there any in which one axis might well be rotated while the other remained in its original position? Carry out several of these rotations and their computations, making a set of drawings to show the points after rotation (all the drawings in which one or the other of the axes have somewhere been shifted will be changed, although individual pairs may not have been rotated; i.e., after the rotation in question 5, all of the graphs involving either F_1 or F_5 will be altered). If another round of rotations were to be carried out, which of the two factor matrices would be used in the multiplication?
7. Compare the communalities of the tests as given by the loadings in question 5, with those indicated by the new factor matrix compiled from the results of question 6. (These should be very nearly identical if all the rotations were orthogonal, otherwise not necessarily.)
8. Describe the method of rotation by setting up trial vectors, instead of shifting by drawings beginning with the unrotated matrix. What are the characteristics of the variable vectors which indicate they are good trial vectors? Mention other methods of starting rotation by jumping to some fairly definite position for the reference vectors.

CHAPTER 13

The Special Problems of Oblique Factors

The pursuit of simple structure—or of any other criterion of rotation which heeds experimental or other evidence of the functional reality of a factor dimension—is likely to cause us to depart from the restricted mathematical formulation in terms of orthogonal factors, as well as from the concept of *positive manifold*, i.e., of variables being loaded only by positive influences of a factor. Orthogonality and positiveness are merely tidiness compulsions in the mathematical, but not scientific, mind. The myth that a positive manifold must be maintained we have already dismissed with the observation that almost any factor is as capable of interfering with a performance as of aiding it. Variables can in any case generally be scored from either end. Even in the realm of abilities we may suffer "the defects of our virtues" and in other fields it is even more clear that traits have disadvantages as well as advantages.

But in this chapter we are particularly concerned with the second point: that the true factor axes obtained by simple structure are likely to be somewhat oblique. Factors in nature do not function in separate universes, but are likely to have some mutual influence and to be somewhat correlated. Indeed to object—on grounds of the mathematical convenience of one's calculations—to correlated factors is to object to the complications of any breadth of causal interconnection in any universe of data.

REASONS FOR AVOIDING EXTREME OBLIQUENESS

However, we should beware of allowing our factors, in the course of the rotation process, to become too highly correlated (oblique), for it seems likely that most factors will manifest decidedly lower inter-

correlations than those found among variables. Experience shows that in the psychological realm, the free pursuit of inherent factor structure rarely finishes with *r*'s of more than 0.4 among the reference vectors, and the majority of *r*'s are below 0.3. Sociological and physiological factorizations have yielded some *r*'s in the region of 0.6 and 0.8, but still the majority lie below 0.3. It would be too early to attempt any serious generalization on the limited number of oblique factorizations yet available, but it is certain, in the psychological realm at least, that for some reason the correlations among factors tend to run lower than those among individual variables.

High correlations among single variables are possible because they can partake to such a high degree of the action of the same factors. They are, indeed, frequently different expressions of one and the same thing, rendered different by slight admixtures of some second influence. In factors, on the other hand, where the causal connection is more likely to be in the nature of an interaction effect between independent (distinct) functional unities, it is natural that the *r*'s would be expected to occupy a lower range. Presumably this would be still more common in higher-order factors, corresponding to influences so much more massive, so much fewer, and so much more remote from one another that they remain unmoved movers, in Plato's phrase, of the factor systems they organize. On the assumption that correlation will exist among first-order factors only when they are not too remote in their realm of operation, Thurstone has referred to an area of correlated factors as a *domain*; but it is possible that we shall find domains to be continuous with one another.

This is not the place to enter into semiphilosophical arguments for the general proposition that the higher the order of the factor system the lower will be the general intercorrelations among them, though such a generalization seems reasonable. However, the argument for avoiding high degrees of obliqueness in the early stages of rotation rests on more practical considerations. To allow them to become highly correlated is to run the risk of (1) confusing or fusing two distinct factors so that one dimension of the space is lost, (2) leaving too much of the factor space unexplored for possible hyperplanes, and (3) losing the guiding influence which one factor can exert upon another.

To expand upon these propositions let us begin by pointing out, with respect to the first, that sometimes the rotator is unknowingly approach-

ing the same hyperplane with two distinct factors. The simple structure which one is intermittently glimpsing and pursuing is really the same for both. If they do not happen to be put on the same graph together for two or three rotations (and with a dozen or more factors this can easily happen) and if one does not happen to notice in the λ matrix a parallelism of pattern among their direction cosines, the first warning of their having become essentially the same factor may come only when one notices that two factors have the same variables in their hyperplanes. Meanwhile, a hyperplane in some other direction must have been completely overlooked. In the successive convergences on denser nebulae of points which have been occurring with respect to the hyperplanes of all factors, the trail to this one missing nebula has somewhere been lost. One must either return to the trial vector taken at the very beginning, with all the waste of time that return entails, or strike out rather blindly with a set of arbitrary direction cosines into that direction of space which up till this point has been most badly neglected.

Third, we must refer to the advantages of mutual guidance when factors are kept not too far from orthogonality. Here one recognizes that the position of the remaining factors is always a rough guide to the most likely position for any given factor. This benefit of guiding one factor by others is best preserved in the first place by considering each round of rotation *as a movement of the whole set of reference vectors*. In an orthogonal matrix, as will be noted from our example in the last chapter, this is forced upon the experimenter, for he does not have the same freedom to shift either all or only some of the factors. It is not true to say that the framework *necessarily* moves as a whole, for parts can be separately unscrewed, but certainly they can only be moved in pairs. When one factor shifts $X°$, the factor it happens to be plotted with has to move $X°$. If there are, say, ten factors, only five independent shifts can be made per round, whereas ten can be made with oblique factors. And since it is as true of the search for simple structure as of some other things that despite general guidance from the group he travels fastest who travels alone, orthogonal rotation is at a disadvantage here.

In oblique rotation it is more necessary to make a deliberate effort to keep in mind that despite freedom to shift individually one has to preserve a feeling for the framework as an integrated whole. Guidance from the whole is particularly evident when one or two

factors have tumbled into their places with very definite hyperplanes and, like pieces in a jigsaw puzzle, help place the more difficult items. For since factors tend not to be *far* from orthogonality, the most likely place to look for a missing hyperplane is in the region roughly at right angles to the soundly established hyperplanes—and it has frequently happened that a reference vector which has obstinately eluded stabilization has been led to a recognizable hyperplane by this method as soon as all its fellow reference vectors have become sufficiently convincing in their hyperplanes to apply it.

One must, however, accept a wide departure from orthogonality if it is very clearly indicated by a hyperplane. Sometimes when none of the hyperplanes are yet very good, such a wide swing occurs in a single leader factor which will drag several neighbors from sterile gropings in the wrong region and bring the whole framework to a good position once the rest are moved roughly orthogonal to it. But until definite leads appear, it is best to keep most factors tolerably orthogonal. Toward the end of rotation, when all hyperplanes have been found but not exactly fixed, orthogonality can be forgotten and excellence of hyperplane fit finally made the *only* goal of rotation.

ROTATION PROCESS RECORDS

Regard for the progress of the rotation as a whole requires the proper keeping of records. The successive λ and V matrices, duly numbered, will naturally be preserved. If one overshoots the best possible hyperplane for a given factor it will then be possible to return to a column in λ and in V which are better than the current ones. As advocated later, it is also advisable to keep a table giving a history of the hyperplane, i.e., of the number of variables appearing in the hyperplane after each rotation, so that the course of improvement can be traced. In this recording one has to use a system which will show how much any individual factor has been rotated, for some will find their hyperplanes ahead of others and may mark time for several shifts on the whole matrix. A left and right superscript is here useful, the right indicating the overall matrix number. Incidentally it generally avoids confusion, especially in making drawings, to go to the labor of carrying forward all λ and V columns even though not all are changed each time.

So far we have not explicitly described how we perceive that a particular pair of factors is converging or diverging in direction,

though it is obvious that if we notice that their direction cosines in the λ columns take on similar or parallel values, the vectors must be folding up together. In the small examples worked out here the angle itself between the vectors can usually be inferred from the shift one made on the previous drawing. But combinations of shifts very soon become too complex for one to infer just what angle still exists between, say, F_2''' and F_3'''. A more systematic and exact way of recording such angles then becomes urgently necessary. In connection with straightening out near-orthogonal factors, at the end of the previous chapter it was pointed out that just as we can calculate the correlation between two test vectors, so we can calculate that between two trial axes by adding the inner products of the loadings (correlations) with the unrotated factors. This practice leads to a new matrix not previously mentioned, which sets out the cosines of the angles among factors. Let us see precisely how this matrix of correlations among the rotated reference vectors, which is usually called C (making a fourth matrix to V_0, λ_n, and V_n), is calculated. An illustration will first be given for the reference vectors used in the example in the last chapter. We shall set out to calculate the C matrix as it stands at the first shift from the orthogonal position (page 206). Let us begin by calculating the degree of obliqueness of F_1' to F_2'. From Table 27 we see that these had direction cosines with the unrotated factors F_1, F_2, and F_3 as shown below in the two columns of Table 29. (In this case we have not turned the columns on their sides, as in earlier instances, to carry out the actual arithmetical processes, since this simplification is no longer necessary.)

TABLE 29.

	F_1'	F_2'	
F_1	$.61\times$	$.53=$	$.3233$
F_2	$.27\times$	$-.78=$	$-.2106$
F_3	$-.75\times$	$.34=$	$-.2550$
	which sum to		$-.1423$

EXAMPLE OF METHOD OF COMPUTATION

The cosine of the angle (the equivalent of a correlation) is thus obtained according to our earlier principle (pages 65 and 207) of working out the inner products. Their addition—to -0.14—shows that the

correlation of F_1' with F_2' in our first attempt at rotation from orthogonality was a perhaps not excessive obliquity, for the angle corresponding to a cosine of 0.14 is a quite small deviation from orthogonality.

To obtain systematically at the end of each round of rotation the new angles among all the rotated factors, it is necessary only to begin with column one, i.e., F_1', and multiply by column two (F_2') and other columns to the right (F_n' etc.) systematically in succession. Then one takes up F_2' and multiplies with all to the right and so on until all possible combinations of factors have been exhausted. These values are entered on a matrix headed *correlations* (or direction cosines) *among reference vectors* or *C*; for our example this is given in Table 30. It will be recognized that C is what is called a symmetric matrix, the upper right and lower left being the same, i.e., the duplication in the upper right is usually omitted.

TABLE 30.

λ Direction cosines of rotated to unrotated factors

	F_1'	F_2'	F_3'
F_1	.61	.53	.76
F_2	.27	−.78	.53
F_3	−.75	.34	.38

C or $\lambda'\lambda$ Direction cosines of rotated factors to one another

	F_1'	F_2'	F_3'
F_1'	1.00	−.14+	.32
F_2'	−.14+	1.00	.12
F_3'	.32	.12	1.00

The reader not familiar with matrix algebra may wonder why the second matrix is headed $\lambda'\lambda$. It is because the process we have described corresponds in matrix rules to *premultiplying a matrix by its transpose*. The transpose of a matrix, λ', is made by changing the columns into rows, so that λ above becomes

TABLE 31. Transpose of a Matrix

		F_1	F_2	F_3
	F_1'	.61	.27	−.75
λ'	F_2'	.53	−.78	.34
	F_3'	.76	.53	.38

By matrix multiplication rules which we have stated earlier, the *columns* of the second matrix (λ) are multiplied by the *rows* of the first (λ') to give the columns of the product. The application of this rule simply leads to our systematically carrying out the operations described above of multiplying the direction cosines of each oblique factor with those of every other to obtain the angles among them. The $\lambda'\lambda$ matrix (or the C matrix as Thurstone's convention labels it) is calculated afresh at each general shift, so that in oblique rotation we typically have dealings with *four* matrices at each round of rotation. The C matrix not only keeps us informed about the general degree of obliquity in the whole framework but also provides the angles between pairs of axes. It is necessary to have the latter to write on each graph, for the interangle constitutes an important aid in deciding which axis to shift, and in which direction, when there is any choice of promising hyperplanes.

OBLIQUE PLOTS: FACTOR AND REFERENCE VECTOR

At this point the student may ask how one plots the graphs when the axes are oblique. This innocent question provokes both a simple and a very complex answer requiring a long digression into the meaning of oblique factors. The first reply, as a practical matter, is that when the obliquity is not great—up to a cosine of 0.5 and occasionally 0.6—one continues to draw the plots as if the reference vectors were orthogonal. This is found to cause very little distortion, and at least the arrangements of points and the results of an angular sweep through them remain closely equivalent on the approximate and the exact graphs. This is best realized by making a practical comparison (for instance for the simple examples above) of these pseudo-orthogonal plots with the true plots obliquely drawn axes—with respect both to the position of the points and the shifts one would have made on them. Since graph paper is not readily obtained for a host of different axis angles, and since the drawer's habits are usually habitually adapted to work with regular graphs, it is much more speedy and practicable to work on orthogonal paper in spite of the slight distortion.

However, with very oblique angles it is more revealing of the general state of affairs to plot with real oblique axes. Various useful devices have been suggested from the use of traveling set squares and parallel rulers to more specialized gadgets, and a few workers

get so skillful with them that they consider it practicable to draw *all* oblique plots in this way.

The mathematical convention for oblique axes is that the projections upon them are carried *not* perpendicularly, as one might imagine and as indicated by the dotted lines upon B' and C' in Diagram 19, but parallel, in respect to projection on one axis, to the other axis, as shown by the continuous lines giving the projections OB and OC in Diagram 22. However, as will be clear after a little more discussion, the values that appear in our V_n matrices after rotation are not really loadings but correlations—they describe the structure, not the pattern, in Holzinger's terminology. The *correlations* of OA with the reference vectors in Diagram 22 are OB' and OC', while the projections or loadings are OB and OC. Our oblique drawings, therefore, can best be made by using ordinary graph paper, reading the horizontal values OC' normally on the paper and pasting the other axis scale at its true angle, running a set square from the OB' values to intercept the verticals from OC' values.

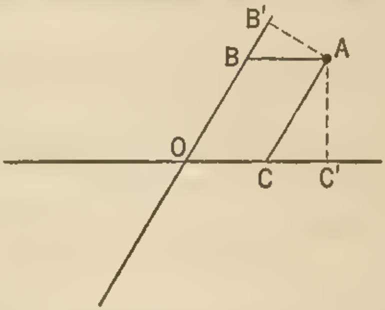


DIAGRAM 22. Manner of Calculating Oblique Projections.

Although we may actually continue to draw orthogonally, the introduction of oblique axes in factor analysis has really brought a need for radical change in our thinking and in our algebraic and quantitative statements. These difficult issues must now be faced. So far we have been accustomed to think of the factor and the reference vector as interchangeable terms, the reference vector at most being thought of as the temporary axis erected at right angles to the hyperplane which, when it comes to rest at the end of the rotation process, we finally call the factor. It is true that as long as we deal with orthogonal factors the terms are synonymous, apart from any such slight, temporary, distinction as that just mentioned, made for convenience in describing rotation. But with oblique factors we have to recognize a *systematic, real, and important difference between the reference vector and the factor vector.*

In oblique factorization we mean by the reference vector that line

through the origin which is perpendicular to the hyperplane, whereas we mean by the factor *the line created by the intersection of the remaining hyperplanes.* The former is familiar and easily understood but the latter will need more illustration. However, let us state here and now that the distinction is important in regard to many calculations and particularly in regard to the use of the specification equation. For the student must accept it as a mathematical truth (if he does not have insight in this realm) that all the calculations which have to do with factors really define the factor as the line of intersection *for all hyperplanes other than that of the factor itself.*

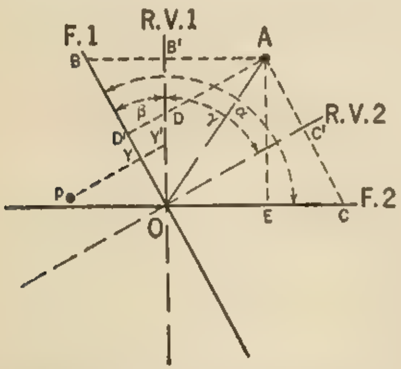


Diagram 23. Projections and Correlations on Factor and Reference Vector.

As a first step in grasping the difference between reference vector and factor we may resort as usual to the simple and concrete example of three-dimensional space. Let the reader visualize the floor and walls at the corner of a room, where the perpendicular to the floor (at the exact corner of intersection) coincides with the line created by the intersection of the walls. Reference vector and intersection vector (factor) are then one. But if one now imagines both walls sloping inward, it is clear that the line of intersection will also slope inward, but the normal to the floor will remain upright. Reference vector and primary factor then have an angular separation and this will be true for all three pairs of reference vectors and factors, i.e., whichever wall one starts with as the hyperplane in question. In short, in this oblique factor situation all the perpendicular reference vectors will diverge from the corresponding lines of intersection created by the other two hyperplanes. (In more than three dimensions the factor is the intersection of *all* the other hyperplanes, for a hyperplane actually has dimensions one less than the whole space.)

To explore this problem further and state in quantitative terms the relations just discussed, it is necessary here to take a two-dimensional drawing, but the reader would do well to consider what

is said in these paragraphs in regard also to a three-dimensional model and to hyperspace, for the full meaning is only evident in that way. In the two-factor instance, set out in Diagram 23, the factor $F1$ corresponding to $RV1$ (reference vector number 1) becomes the intersection with the plane of the paper of the hyperplane connected with reference vector number 2 ($RV2$). Similarly if we want to locate $F2$, we look for the line of intersection of the hyperplane which is attached to and fixed by $RV1$. Let us now take a test vector OA and compare its projections on the alternative systems[1] presented by the F's and the RV's.

TRANSLATION FROM REFERENCE VECTOR TO FACTOR

Now the *loadings* of the test vector OA on factors 1 and 2 are respectively OB and OC, as shown by the proper mathematical convention for oblique projections (carrying each projection line parallel to the other axis instead of perpendicular to the axis on which it falls). But the *correlations* of OA with these factors remain as given by the proper cosine convention, namely OD' and OE. However, we are not for the moment going to do anything about these correlations with the factor, which are not given us directly in any data we have. What we *do* have in our rotation data is *a set of relations to the RV*'s and what we need to notice immediately from Diagram 23 is that the *correlations* of OA with the RV have simple relations to its oblique *projections* (loadings by definition) on the F.

As given by our cosine rule the correlations of OA with $RV1$ and $RV2$ are OB' and OC' (keeping OA of unit length for simplicity).

[1] Holzinger (71, 74), followed by Harris (67, 68), has suggested a different nomenclature for these alternative systems which gives greater importance to the reference vector system, in accordance with an argument independently put forward by the present writer (24). But the present writer is inclined to recant from his position (except when confronted with conservative extremists who can see no use whatever for the reference vector system!) and has adopted above the usual nomenclature which calls the first and older system an F (factor) system and the latter an RV (reference vector) system, implying that the former is the more widely useful equivalent of the factors of the algebraic and logical analysis.

However, Harris and Holzinger's terminology may well be tried for its usefulness. What we call a factor they call a primary factor, and what we call reference vectors they call a simple factor. They also raise to equivalent factor status two other concepts based on clusters which we have rejected earlier as unsatisfactory guides for factor rotation. When an axis is put through a cluster, they call it a cluster factor, and since there is also here the F and RV alternative, here the equivalent is called a normal factor. The two last are clear enough concepts though we do not need to use them here.

Now it will readily be seen that though OB (on $F1$) and OB' (on $RV1$) are never equal (so long as $F1$ and $RV1$ are distinct), they remain (for all tests) in a fixed ratio for any given reference vector and its factor. In fact, if the angle between the $RV1$ and the $F1$ is β, as drawn in Diagram 23, the *correlation* of OA with the $RV1$ equals its *loading* on $F1$ multipled by cosine β, and the same for any other variable. Thus a simple numerical translation is always possible between RV and F with respect to either correlations or loadings.

Incidentally, the angle immediately used in these translations is obviously β, not α, the angle between the factors which we normally get by calculation in the C_F matrix. But the angular relations are not complex, the angle between the factors, α, being in this case the supplement to the angle between the reference vectors, γ, while the latter is the complement of β, the required angle. Consequently if we want to express the above proportionality immediately in terms of the angles between the factors, we can say

$$\frac{\text{Correlation of variable with } RV}{\text{Loading of variable with } F} = \text{Sine of angle, } \alpha \text{, between factors} \tag{25}$$

Similarly, if we wished to deal with the transformation from the *loading* on the reference vector to the *correlation* with the factor (note this is *not* the converse of the above), we follow the projection line ADD' (parallel to $RV2$). The values given, OD and OD', likewise remain in a fixed proportion—the reciprocal of the sine of angle between factors. Indeed, in general, there is complete reciprocal relationship between factors and reference vectors, mediated by a coefficient corresponding to the sine of the angle between factors.

The distinction between reference vector and factor thus turns out to be closely associated with the difference between loading and correlation. Although we have kept projection and correlation as distinct concepts, we have not needed to treat them differently so long as we dealt with orthogonal factors. The difference, however, has been systematized by Holzinger and Harman (**71**) in the concepts of factor *pattern* and factor *structure*. (The latter term, we may remind the reader again, is not to be confused with correlation configuration, sometimes called correlation structure, having to do, respectively, with the fixed relations of the vectors in space and relation of factors to test vector distribution. To avoid confusion, we have

throughout used *factor resolution* for the latter.) The specification equation, or rather, the whole set of specification equations and, therefore, the factor matrix, are defined as the *factor pattern*. It shows the composition of the tests in terms of *loadings* or projections on the factor and is used in predicting a test performance from factor endowment. The *factor structure*, on the other hand, shows the *correlations* between the tests and the factors. The latter is used in telling us what to add together in order to estimate the factors, but it is the former we need when we want to make a prediction of the performance of an individual from a combination of factors.

In using oblique factors no difficulty ensues if we remember to employ the formulations appropriate to the system of values we are using and if we remember the special advantages of each for particular computations or drawings. For example, graphical rotations away from orthogonal positions and in search of simple structure necessarily deal with reference vectors, for these lie normal (at right angles) to the visible nebulae of hyperplanes obtained by our plots. The factor resolution we obtain by all oblique rotation methods is therefore first a simplification of factors *in terms of their reference vector system.*

Let us now see what happens when our factor resolution reaches simple structure by rotation of the *RV*'s. In general we shall have been using simplified, approximate drawings making the RV's orthogonal which may confuse our perception of what is actually happening. But essentially we have reached a position where a lot of variables, like p in Diagram 23, have zero (or practically zero) *perpendicular projection* upon the reference vector (*RV*1 in this case). This means that they have *no correlation with the RV* and therefore have zero entries in the V_n matrix. The latter has been obtained at each rotation as a transformation of the correlations with the original unrotated factors or reference vectors (which are identical in the unrotated situation) into correlations with the rotated reference vectors.

If such points have no correlation with the *RV*'s, they will have no loadings on the factors, as Diagram 23 reminds us, so we have achieved simple structure at the same time in terms of the *loadings* on the factors—the F matrix. On the other hand, as a glance at the point p will show, such variables lying in the graphical hyperplane may have a substantial *correlation* (OY in this case) with the factor and proportionately large *loading* (OY') on the reference vector. The

factor *structure,* in short, will not show simple structure. Simple structure belongs to the *RV* structure and to the factor pattern. Its absence from the factor structure need not bother us, for our more important computations are with factors. But this anomaly should stimulate us as a practical and theoretical issue at least to weigh the advisability of adapting for general use the converse system, i.e., the reference vector system, instead.[2]

RELATIVE ADVANTAGES OF F'S AND RV'S

The final decision will have to be guided by such facts as that the present system gives relatively simple prediction of test performance from factor endowments, since there are many loadings that are zero. But the estimation of factors from tests, on the other hand, is not so simple, for there are practically no tests with zero *correlations* with the factor. The converse pros and cons hold if we deal with reference vectors. However, the factor system, with the first advantage, is probably the greater convenience, since in practice we should usually need to estimate a person's factor endowments but once; whereas we need to apply them many times in the situations represented by the numerous specification equations. Moreover, we do not normally go to the trouble of using *all* tests with significant correlations with a factor when we want to estimate a person's factor endowment—we use the top few, which would be no more numerous with factors than with reference vectors.

The only practical drawback to the expression of all final results in terms of factors rather than reference vectors is that the rotation process is carried out in terms of correlations with the reference vectors; so that when simple structure is finally reached, a transformation has to be made converting the correlations with the *RV*'s, as given in rotated factor matrix V_n, into the loadings on the factors (which we may call matrix F_n). But since, for any one factor, the r's with the *RV* are simply proportional to the loadings on the *F*, the order and relative importance of all the variables on a given factor

[2] At one point in the history of this debate (24), as stated in the last footnote, the present writer has argued the merits of the reference vector system and criticized psychologists bound by the formal, approved mathematical nomenclature of factors for not at least opening to discussion the question of whether the reference vectors might not better be called factors. The reader is now in a better position to debate these alternatives to himself, but to the writer it seems best to adopt the compromise of special uses for each as indicated in these paragraphs, while still reserving the term *factor*—or at least *primary factor*—for only one.

will not be changed by the transformation. If the aim of a particular research is to discover the *nature* of the factors at work in a given area, therefore, there is no point in making the transformation; for one is interested only in knowing the highly loaded variables by which the factor is characterized and the nature of the variables which fall in the hyperplane and which define it by a ground effect.

True, even for this purpose there is some slight drawback in that one does not get the exact relative size of factors, i.e., their average contribution to the variance of all the tests; nor can one decide with certainty whether a particular variable is more highly associated with one factor than another when both influence it strongly, since the ratio by which r's are multiplied to transform them to loadings will differ with different factors, according to the cosine of the angle between RV and F. And the latter information as to relative loading, e.g., as to whether a test of, say, mechanical ability at the top of two factors F_x and F_y is higher in F_x than F_y, does contribute somewhat to the interpretation of the factors. But with the slight degree of obliqueness commonly found among factors, this boosting of the over-all variance of some of them is not enough to cause real distortion.

Consequently, the present writer has urged (**24**) that if the object of the research is as stated above (and at present most factorizations are exploratory), i.e., if it is to determine the nature of the factor and is concerned only with a rough estimate of its relative contribution to general test variance and not at all with predictions for particular individuals via the specification equation, the results should be presented without transformation. This convention of presenting (a) the loadings actually in terms of the r's on the RV's to which they are proportional and (b) the factor intercorrelations actually as *reference vector* intercorrelations has been followed satisfactorily by the present writer in some dozen factorizations and by many other researchers, without any misunderstanding and with economy of research time.

However, in any *ultimate* discussion of the nature of a factor it is desirable to make the transformation to factor values, because a proper appreciation of its meaning depends not only on the loading pattern but also on inspecting the correlations which determine the factor estimates, the loadings which show the true relative variance of factors as a whole and, especially, the inverse of the $\lambda'\lambda$ matrix which gives the true correlations among factors, indicating the nature of second order factors among them.

COMPUTATION OF FACTOR LOADINGS FROM RV VALUES

The argument for leaving the results of exploratory researches in the *RV* system will be better appreciated by the student when he discovers from experience how much computation is implied by the simple statement in the last sentence that the inverse of a matrix is to be calculated! We must now face this task in setting out the complete computation for changing from reference vectors to factor values at the end of the reference vector rotation. Our stock in trade so far has been an unrotated factor matrix, V_0; a transformation matrix, λ; a rotated factor matrix, V_n (actually an *RV* matrix); and an interfactor or *RV* direction matrix, $C = \lambda'\lambda$. The aim is now to get in addition a true rotated factor matrix, F_n; a factor transformation matrix, λ_F by which to obtain the former from V_0; and a true interfactor matrix, $C_F = \lambda'_F\lambda_F$. For clarity in what follows, we shall attach the subscript, R to make the usual C, λ, λ' and V^n appear as C_R; λ_R; λ'_R, and V_{Rn} respectively, indicating that they refer to our reference vector system and thereby differentiating them from C_F, λ_F, λ'_F, and V_{Fn}.

Before proceeding further let us recall a few definitions which we shall need. *The principal diagonal* of a matrix is the diagonal starting at the upper left-hand corner of the matrix and extending down to the lower right-hand corner. The *diagonal element* of any element to the right of the principal diagonal is that element of this diagonal which lies in the same *row* as the given element. The diagonal element of any element below the principal diagonal is that element of this diagonal which lies in the same *column* as the given element. An *identity matrix* is a matrix with 0's everywhere except along its principal diagonal and the elements along the principal diagonal are all 1's. The symbol I is usually used to signify an identity matrix. The matrix C^{-1} is said to be the *inverse* of the matrix C if and only if the relationship $C \cdot C^{-1} = 1$ holds between them.

From the two-dimensional model above (Diagram 23) we might generalize that

$$\text{Factor structure} = (\text{Reference vector pattern})\ (\sin\beta)$$

or

$$\text{Factor pattern} = (\text{Reference vector structure})\ (\cos\beta)$$

where β is the angle between RV and F. The solution of our transformation from V_{Rn} to V_{Fn} correspondingly hangs upon finding a diagonal matrix which we will call D, giving the cosines (or sines) between the RV's and the corresponding F's. By saying a diagonal matrix we mean that it will have values other than zero only along the principal diagonal, each connecting one reference vector with one factor.[3] It can be shown that this D matrix is related to the alternative transformation matrices as follows:

$$\lambda_F' \cdot \lambda_R = D$$

The process of determining D (and therefore λ_F, C_F, and V_{Fn}, or F_n, as it is sometimes written) is as follows:

First one finds the inverse of the C_R matrix. This is the most bothersome part of the undertaking. Although, as explained on page 182, an 8 by 8 triangular matrix may have its inverse calculated by a skilled computer in about an hour, the computation for a full symmetrical matrix of the same size would take at least four hours, and the matrices in some of the factor problems discussed here may need one or two days for obtaining the inverse. Calculating the inverse of the 5 by 5 matrix given in the example below took about an hour and a half.

In addition to the standard procedures of matrix inverse calculation provided in matrix algebra textbooks, a number of quicker methods adapted to special purposes have been published from time to time. A recent survey of some is given in Dwyer's *Linear Computations* (**46a**). A quick method for the inverse of a *triangular* matrix is given in the multigroup extraction method above in Chapter 13. A good general method for both symmetrical and nonsymmetrical matrices is that of Tucker, adapted and set out by Thurstone (**126**, page 46). A method quicker for small matrices, simple numbers, and sophisticated computers is that of Andree (**1**). The method of Fruchter has already

[3] As Harris and Knoell compactly state (68), "D represents only the operation of normalizing either the rows or the columns." The above equation when rearranged means that "the transformation matrix of one is equivalent to the transpose of the normalized inverse of the transformation matrix of the other." Further rearrangement of the above will show that we can get not only one transformation from the other but also the angles among the factors from the angles among the reference vectors, thus:

$$\lambda_R \lambda_R' = (\lambda_F'^{-1} D)(\lambda_F'^{-1} D)'$$

$$\text{and } \lambda_F' \lambda_F = (D \lambda_R^{-1})'(D \lambda_R^{-1})$$

been mentioned (**56**). The latter is similar to that advocated here, adapted by Saunders from Crout (**40**). This is the most time-saving, with existing computing aids, fair-sized matrices, and reasonably skilled statistical clerks, known to the present writer.

Commonly, and in the present task of shifting from RV pattern to Factor pattern, the computer has to deal with a symmetrical matrix (in this case the C matrix), and therefore in the example below we have used a symmetrical matrix. However, for general usefulness the following account of procedure is set out for the general, non-symmetric case with indications of what steps can be shortened when the matrix is symmetric.

The work consists in finding two matrices: (1) an auxiliary matrix, and (2) the final matrix, the inverse we are seeking. This method is particularly applicable when a computing machine is available, for each element can be obtained by a continuous machine operation—an algebraic sum of products with, in some cases, a final division.

Let us assume we have completed rotation on a five-factor problem and arrived at a C_R interreference vector matrix as shown in Table 32. (Henceforth we put RV instead of F at the heads of columns in virtue of the now recognized distinction.)

TABLE 32.

$(C_R = \lambda_R' \lambda_R)$

	RV_1	RV_2	RV_3	RV_4	RV_5
RV_1	1.00	.10	−.20	.70	.20
RV_2	.10	1.00	.30	−.40	−.20
RV_3	−.20	.30	1.00	.60	−.30
RV_4	.70	−.40	.60	1.00	.50
RV_5	.20	−.20	−.30	.50	1.00

The sequence of steps is as follows:

1. Extend the given matrix by placing an identity matrix of the same order alongside it on the right. Thus, for our example, since it is of order 5, we append the 5 by 5 identity matrix to it and obtain

$$(1)\quad \left|\begin{array}{rrrrr:rrrrr} 1.00 & .10 & -.20 & .70 & .20 & 1.00 & 0 & 0 & 0 & 0 \\ .10 & 1.00 & .30 & -.40 & -.20 & 0 & 1.00 & 0 & 0 & 0 \\ -.20 & .30 & 1.00 & .60 & -.30 & 0 & 0 & 1.00 & 0 & 0 \\ .70 & -.40 & .60 & 1.00 & .50 & 0 & 0 & 0 & 1.00 & 0 \\ .20 & -.20 & -.30 & .50 & 1.00 & 0 & 0 & 0 & 0 & 1.00 \end{array}\right|$$

2. To obtain the auxiliary matrix, which will have the same number of rows and columns as (1), we proceed as follows:

a. Copy the first column of (1) into the first column of the auxiliary matrix (2). They are identical.

b. The remaining elements in the first row of (2) are obtained by dividing the corresponding element in (1) by the first element in the row. In our example the first row of our auxiliary matrix turns out to be the same as the first row of (1) because the first element is 1.00 and division by 1.00 does not change the other elements; but this, of course, would not be the case where the first element was different from 1.00.

c. In computing the remainder of the auxiliary matrix one alternates between columns and rows and finds, first, those elements of column 2 still undetermined; secondly, those elements of row 2 still undetermined; thirdly, those elements of column 3 still undetermined; fourthly, those elements of row 3 still undetermined, etc., until the matrix is completed.

To compute these elements one uses the following rules:

(A) If the element lies on or below the principal diagonal, take the corresponding element in the original matrix (1) and subtract from it the sum of the products of the elements (in the auxiliary matrix) in its row with the corresponding element in its column, using, of course, *only elements which have been previously computed.*

(B) If the element lies above the principal diagonal, proceed exactly as in rule (A) but, in addition, divide the result by the diagonal element (in the auxiliary matrix) of the element being sought. In doing the calculations a helpful device is to encircle or underscore (shown by italics below) the diagonal elements to make them stand out.

For our example the auxiliary matrix turns out to be the following:

$$
(2)\left|\begin{array}{rrrrr|rrrrr}
\mathit{1.00} & .10 & -.20 & .70 & .20 & 1.00 & 0 & 0 & 0 & 0 \\
.10 & \mathit{.99} & .3232 & -.4747 & -.2222 & -.1010 & 1.0101 & 0 & 0 & 0 \\
-.20 & .32 & \mathit{.8566} & 1.0412 & -.2205 & .2712 & -.3773 & 1.1674 & 0 & 0 \\
.70 & -.47 & .8919 & \mathit{-.6418} & -.7046 & 1.5416 & -1.2641 & 1.6223 & -1.5581 & 0 \\
.20 & -.22 & -.1889 & .4523 & \mathit{1.1882} & -.7307 & .6082 & -.4319 & .5931 & .8416
\end{array}\right|
$$

Letting R and C stand for row and column respectively, some sample calculations for the example are as follows:

R1C4: $.70 \div 1.00 = .70$

R2C2: $1.00 - (.10 \times .10) = .99$

R5C2: $-.20-(.20\times.10)=-.22$
R2C5: $[-.20-(.10\times.20)]\div.99=-.2222$
R2C8: $[0-(0\times.10)]\div.99=0$
R3C3: $1.0-(.20\times.20+.32\times.3232)=.8566$
R4C3: $.60-(-.20\times.70-.47\times.3232)=.8919$
R3C6: $[0-(-.20\times1.00-.32\times.1010)]\div8566=.2712$
R4C5: $[.50-(.70\times.20+.47\times.2222-.8919\times.2205)]\div(-.6418)$
$=-.7046$
R5C8: $[0-(.20\times0-.22\times0-.1889\times1.1674+.4523\times1.6223)]$
$.\div1.1882=-.4319$

When the original matrix is symmetric one can greatly shorten the calculations for the elements in the triangular array above the principal diagonal and to the left of the dotted line by applying the following rule: To obtain any element in the aforementioned triangular array take its symmetrically opposite element (in the auxiliary matrix) and divide it by its diagonal element (in the auxiliary matrix). Indeed, if one is using a computing machine one would merely keep the answer in the machine when computing an element below the principal diagonal and, after recording it in the proper space below the diagonal, would then divide by its diagonal element and record the answer in the symmetrically opposite position. Thus, we have, for example,

R2C4: $-.47/.99=-.4747$
R3C4: $.8919/.8566=1.0412$
R4C5: $.4523/-.6418=-.7046$

3. In obtaining the final matrix we shall work with two sections of the auxiliary matrix: (a) with that part of the matrix to the right of the dotted line and we shall call this matrix A; and (b) with the triangular array of elements above the principal diagonal (i.e., to the right of it) but to the left of the dotted line which we shall call T. The final matrix may be considered as proceeding from matrix A.

The sequence of steps is as follows:

a. The bottom row of the final matrix is identical with the bottom row of A.

b. In determining the remaining elements one always works from the bottom to the top of the column, taking first column 1, then column 2, next column 3, etc. until all columns are complete. The rule for calculating these elements is as follows: Take the corresponding element in matrix A and subtract from it the sum of the products of the elements already calculated in its column in the final matrix (starting at the bottom) with their corresponding elements in its

row in T (starting at the dotted line and working toward the diagonal element).

The final matrix of the example turns out to be the following:

$$(3)\ \left|\begin{array}{rrrrr} .1823 & .5340 & -.9589 & 1.0267 & -.7306 \\ .5339 & .5460 & .6267 & -.8354 & .6081 \\ -.9589 & .6268 & -.3001 & 1.3180 & -.4319 \\ 1.0267 & -.8356 & 1.3180 & -1.1402 & .5930 \\ -.7307 & .6082 & -.4319 & .5931 & .8416 \end{array}\right| = C_R^{-1} = (\lambda_R'\lambda_R)^{-1}$$

Some sample calculations would be

R4C1: $1.5416-(.7307\times .7046)=1.0267$

R3C1: $.2712-(.7307\times .2205+1.0267\times 1.0412)=-.9589$

R1C1: $1.00-(-.7307\times .20+1.0267\times .70+.9589\times .20+ .5339\times .10)=.1823$

R4C2: $-1.2641+.6082\times .7046=-.8356$

R2C2: $1.0101-(-.6082\times .2222+.8356\times .4747+.6268\times .3232) =.5460$

R3C3: $1.1674-(.4319\times .2205+1.3180\times 1.0412)=-.3001$

R4C4: $-1.5581-(-.5931\times .7046)=-1.1402$

R1C5: $0-(.8416\times .20+.5930\times .70+.4319\times .20+.6081\times .10) =-.7306$

The final matrix (3) is the inverse, C_R^{-1}, we were seeking.

If desired, one can carry along a check column to give a continuous check on calculations. This is done as follows:

(A) Add all rows in (1) and write the sums as an additional column to the right of (1).

(B) In calculating both (2) and (3) treat the extra column the same as any of the columns to the right of the dotted line.

(C) In the auxiliary matrix any element in the check column should be equal to one plus the sum of those elements in its row which lie to the right of the principal diagonal.

(D) In the final matrix an element in the check column should equal one plus the sum of the other elements in its row.

For our example the check columns for (1), (2), and (3), respectively, are as follows:

2.80	2.80	1.0535
1.80	1.5353	2.4793
2.40	2.8820	1.2539
3.40	.6371	1.9619
2.20	1.8803	1.8803

An alternate check would be to take the product of the original matrix, C_R, given in Table 32 with its inverse, C_R^{-1}, given in (3) and see if it turns out to be the identity matrix.

In order that the reader might better understand the process of finding an inverse, we used a 5-by-5 matrix in the above example instead of the simpler 3-by-3 matrix we had been working with. For the remainder of the chapter we return to our former example and follow through with it. The C_R matrix for our example was given in Table 30 (page 215). Following the method outlined above, we find the inverse of this matrix to be

$$C_R^{-1} = (\lambda'_R\lambda_R)^{-1} = \begin{vmatrix} 1.15 & .21 & -.40 \\ .21 & 1.05 & -.20 \\ -.40 & -.20 & 1.15 \end{vmatrix}$$

We can now proceed to calculate the intercorrelations of the factors, i.e., the matrix C_F or $\lambda'_F\lambda_F$, from the equation

$$C_F = D\, C_R^{-1}\, D$$

where D is a diagonal matrix so chosen as to make the elements of C_F along its main diagonal equal to one. This is accomplished by choosing for each diagonal element of D the reciprocal of the square root of the corresponding element in C_R^{-1}.

Thus, taking the diagonals from C_R^{-1} in our example, we find D to be

$$D = \begin{vmatrix} \frac{1}{\sqrt{1.15}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{1.05}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{1.15}} \end{vmatrix} = \begin{vmatrix} .93 & 0 & 0 \\ 0 & .98 & 0 \\ 0 & 0 & .93 \end{vmatrix}$$

We can thus find the products

$$D \cdot C_R^{-1} = \begin{vmatrix} 1.07 & .20 & -.37 \\ .20 & 1.03 & -.19 \\ -.37 & -.19 & 1.07 \end{vmatrix}$$

and

$$(D \cdot C_R^{-1})D = \begin{vmatrix} 1.00 & .19 & -.34 \\ .19 & 1.00 & -.18 \\ -.34 & -.18 & 1.00 \end{vmatrix} = C_F$$

We have already seen that $\lambda_F \cdot \lambda_R = D$. It can be shown by matrix algebra that $(\lambda'_R \lambda_R)^{-1} = \lambda_R^{-1} \lambda'^{-1}_R$. Using these values we obtain the following expansion:

$$DC_R^{-1} = D(\lambda'_R \lambda_R)^{-1} = D\lambda_R^{-1}\lambda'^{-1}_R = (\lambda_F \lambda_R)\lambda_R^{-1}\lambda'^{-1}_R \qquad \text{(A)}$$

By definition of an inverse, $\lambda_R \lambda_R^{-1} = I$, the identity matrix, whose use as a multiplier among matrices is the equivalent of 1 among ordinary numbers. Hence, in the last term of (A) the two inside elements become the identity matrix and (A) reduces to

$$DC_R^{-1} = \lambda_F I \lambda'^{-1}_R = \lambda_F \lambda'^{-1}_R \qquad \text{(B)}$$

Taking the first and last terms of (B), let us now multiply each of them by λ'_R on the right hand side, yielding

$$DC_R^{-1}\lambda'_R = \lambda_F \lambda'^{-1}_R \lambda'_R \qquad \text{(C)}$$

But, since $\lambda'^{-1}_R \lambda'_R = I$, the right-hand side of (C) reduces to λ_F and we have

$$DC_R^{-1}\lambda'_R = \lambda_F$$

Let us now proceed to apply these equations to our present example. We found above that

$$DC_R^{-1} = \begin{vmatrix} 1.07 & .20 & -.37 \\ .20 & 1.03 & -.19 \\ -.37 & -.19 & 1.07 \end{vmatrix}$$

Multiplying λ'_R (given in Table 31, page 215) by this matrix, we have:

$$\lambda_F = (DC_R^{-1})\lambda'_R = \begin{vmatrix} 1.07 & .20 & -.37 \\ .20 & 1.03 & -.19 \\ -.37 & -.19 & 1.07 \end{vmatrix} \cdot \begin{vmatrix} .61 & .27 & -.75 \\ .53 & -.78 & .34 \\ .76 & .53 & .38 \end{vmatrix}$$

or

$$\lambda_F = \begin{vmatrix} .48 & -.06 & -.88 \\ .52 & -.85 & .13 \\ .49 & .61 & .62 \end{vmatrix}$$

To check our result, we may form the product $\lambda_F \lambda_R$ and obtain D (with allowance for rounding error).

Check:

$$\lambda_F \lambda_R = \begin{vmatrix} .94 & .00 & .00 \\ .00 & .98 & .00 \\ .00 & .00 & .93 \end{vmatrix}$$

A further check shows us that $\lambda_F \lambda'_F = C_F$, again with allowance for rounding error.

From λ_F, along with the original, unrotated matrix V_0, the true pattern of factor loadings can now be obtained by the usual step:

$$F_n \text{ or } V_F = V_0 \lambda F$$

Since this calculation—principally the computing of a matrix inverse—can be very time consuming when, say, ten or more factors are involved, it is recommended, as stated above, that the solution be left in terms of reference vectors when circumstances permit. These matrices having to do with reference vectors should be clearly labeled as such, and sufficient data published to enable whoever wishes to proceed to the factor loadings, the factor structure, and the extraction of second order factors.

The publication of inadequate data—and the inadequacy has sometimes extended even to publishing rotated factor matrices without any indication of the angles among the vectors or without the V_0 matrices from which they are derived—is strongly to be deprecated, whether it springs from the slovenliness of writers or the parsimony of editors. An adequate statement of data is one which permits checks to be made and the calculation of any matrices not actually presented. If space, computing time, etc. permit, it is a welcome luxury to present a factor matrix, but R (correlation), V_o, λ_R, C_R, and V_{Rn} matrices suffice. Indeed a polished presentation of the complete factor solution is vain if the time for it has been gained at the expense of abbreviating, however little, the careful rotation of the reference vectors themselves for simple structure. For it is upon the angles found for these hyperplanes that all further transformations depend.

Questions and Exercises

1. State the arguments for the closer correspondence of oblique than orthogonal factors to the structure in nature. Why in the actual practice of rotation should one avoid extreme obliqueness?
2. In a set of rotations where orthogonality of pairs is preserved in each rotation, is it true that all axes will be mutually perpendicular at the end of the complete set of rotations? Why?
3. Describe the meaning and mode of calculation of the C matrix. Why are the entries on the main diagonal of the C matrix in Table 30 all equal to 1? Will this always be the case in such a product matrix?

4. What is meant by factor structure and factor pattern? Discuss with a diagram for a two-dimensional problem, the relation of loadings and correlations of test vectors with respect to (a) factors and (b) reference vectors. Under what conditions will loadings and correlations coincide?
5. What is meant by (a) the transpose of a matrix and (b) the inverse of a matrix?
6. Describe, either in words or in matrix notation, the steps necessary (omitting any details of computation) for arriving at the matrices of a factor solution from those of a reference vector solution.
7. What types of research may reasonably leave their solutions in reference vector terms? Discuss the restrictions on inference from such formulations.
8. Find the inverses of the following matrices:

a. $M_1 = \begin{pmatrix} 2 & 3 \\ 5 & 4 \end{pmatrix}$; b. $M_2 = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \end{pmatrix}$; c. $M_3 = \begin{pmatrix} 3 & 1 & -4 \\ 5 & 2 & 2 \\ 1 & -1 & -3 \end{pmatrix}$

d. $M_4 = \begin{pmatrix} 1 & 0 & 0 & x_0 \\ 0 & 1 & 0 & y_0 \\ 0 & 0 & 1 & z_0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

CHAPTER 14

General Techniques and Criteria of Factor Resolution

Anyone who has surveyed the results of factor analysis in many fields since its dim beginnings fifty years ago or even since its lustier adolescence of twenty years ago must admit that about half of the published analyses are abortive, inconclusive, and sometimes positively misleading. These failures arise from a variety of faults in design to which we shall give attention in due course, such as lack of alertness to sampling of variables, incompleteness of factorization, etc; but the one single cause which outtops all others is failure in the rotation process.

CAUSES OF ROTATION FAILURE

Rotation *may* fail through actual errors or wrong theoretical concepts about the goals of rotation, but a considerable proportion of the spurious solutions arise from using a principle, namely, simple structure, that is valid in most circumstances but which is applied with defective skill or persistence. To take proper cognizance of a very human consideration, we must realize that the trial-and-error search for simple structure with present methods and machines is a very exhausting and sometimes exasperating procedure. An experimenter with less than a true explorer's determination is all too prone to give up just when he is getting in sight of his goal. For the rotation of a ten or twelve factor matrix with forty to sixty variables is likely to take one person's full time for three months, or up to five or six if the interfactor angles are to be well determined, since about two dozen *general* rotations may be necessary, each requiring a week.

It is not surprising, therefore, that many researches get terminated as soon as some degree of definition of hyperplane has appeared, yet

this is precisely the point where a fresh zeal to attain the goal of a really neat hyperplane fit should be generated. For the three or four general rotations that are made at this phase are usually the most rewarding of all, and the clear-cut, imperative hyperplanes which now appear are likely to make those which have previously been accepted seem very vague. Moreover the last few shifts are apt to cause sharp changes in angle in a few factors, changes which may alter their psychological meaning considerably.

It is our purpose in this chapter, therefore, to examine possibilities of improving the rotation process by substituting analytic or semi-analytic techniques for the groping of trial and error, or by bringing certain efficiency aids to the trial-and-error process itself. To be able to judge the effectiveness of these various methods it is necessary to be able to define more exactly when the goal of simple structure is reached; and we shall accordingly first consider possible criteria to test a good simple structure.

GOODNESS OF HYPERPLANE FIT

The simplest and most widely used procedure is to examine the goodness of fit of each single supposed hyperplane. Later, from the expressions for the single hyperplanes one may obtain an expression for the goodness of the solution as a whole. Common experience has led researchers to consider that with the typical populations and tests, a loading within ± 0.05 or ± 0.10 can be regarded as essentially a zero loading, i.e., as lying in the hyperplane. It is a valuable procedure to count the numbers within these two limits (giving two separate but not entirely independent indexes) and enter them at the bottom of the V matrix columns for each successive rotation. A separate, auxiliary table can also advantageously be set out and labeled *history of the hyperplanes* on which the numbers of variables in the supposed hyperplanes is recorded, row after row for the successive rotations with respect to each factor.

This same record sheet should have space for notes on a variety of matters which, as the present discussion will indicate, need to be organized in the art of rotation. For example, in each row indicating a new general, overall rotation one should mark the few factors[1] which

[1] It will be understood that in *general* contexts the term *factor* will continue to be used in a broad, generic sense for either reference vector or true factor, according to which system one happens to be working in.

are carried forward *without* a shift. Secondly, it helps later decisions to record on what other *RV*'s the given *RV* was shifted each time, which can be done by writing in the history table the reference numbers of the *RV*'s involved, immediately above the hyperplane numbers which they helped to produce. Also, one should indicate by a question mark the rotation at which the drawing of a given factor presented two almost equally promising hyperplanes and at which one path at the crossroads was followed; for if it ends unsatisfactorily, one can then turn back readily to what was probably the right alternative. Again, if two factors later fold up, one into another, one may find that an earlier, independent stage of one, which had a good hyperplane, can be taken up again as a substitute for some factor which has petered out in an unstructured wilderness, having not even the ghost of a hyperplane. On this *history* record, the first evidence of a good hyperplane being reached is the observation that the number of points in the hyperplane has climbed to a plateau and stayed there. Normally one should have on the record at least four or five rotations of a given factor without further improvement of the hyperplane (and sometimes with temporary intervening loss) before assuming that the best position has been found. This record will also enable one, if the best position has been overshot, to go back and pick up the direction cosines corresponding to the best position.

After the plateau test, the next most important indicator is the *percentage* of variables lying in the supposed hyperplane. This connects with Thurstone's criterion of at least one zero loading for each test and for each factor. In the writer's opinion, this statement has the value of a general slogan, but, taken literally, is too rigid and asks too much in one respect and too little in another. However it has the virtue of reminding one that the fit of a single factor cannot be properly judged without regard to the simple structure picture in the whole factor matrix.

WIDTH OF HYPERPLANE

The question of the acceptable width of the hyperplane, passed over with a rule of thumb as regards the plateau test, where only *relative* goodness of hyperplane was involved, is more crucial and must be examined more closely when we consider this new criterion of simple structure using the absolute amount or area of hyperplane. The present writer can report as a matter of empirical evidence that

he has sometimes counted loadings through ± 0.05, sometimes through ± 0.10, and sometimes through ± 0.20; but generally has kept records of two ranges, chiefly ± 0.05 and ± 0.10. If all is well, they show a marked tendency to vary together, and if occasionally a shift brings an improvement on one and a loss on the other, one can best decide its goodness by weighting the number found within ± 0.05 about twice the number within ± 0.10.

This practice has worked well enough in a wide variety of problems, but it is obvious that eventually a more sensitive and adapted check must be employed when researches pass the exploratory stage and that such a criterion will take account of the standard error of the factor loadings. To anticipate the discussions of Chapter 17, we may recognize that the errors of factor loading arise broadly from errors of measurement (low reliability of the original r's) and errors of sampling. Consequently the hyperplane width accepted must be a function of these. The loading error, in the course of rotation, will spread itself evenly among the factors, so that the same criterion can be taken for all. But until Chapter 17 we shall make no attempt to discuss the estimation of these loading errors since it is a debatable matter. Meanwhile, as a more approximate guide to what shall be counted in the hyperplane we can take the value which will already be worked out in deciding when to stop factoring, i.e., the figure used for deciding what residuals in the factor extraction should be considered negligible (page 296). For we can take two or three times the sigma of these residuals as the limiting value for the hyperplane width. Whatever value is taken, it is helpful to take two counts—one for a broader and one for a narrower hyperplane. With populations of 200 or more and variables of reliability around 0.7 to 0.9, as indicated above, the writer has repeatedly used the ranges of ± 0.05 and ± 0.10, at least without any conspicuous failure or anomaly.

AREA OF HYPERPLANE

When this criterion for being counted in the hyperplane has been settled, we can employ the second test of simple structure mentioned above, namely, the hyperplane area, by which we mean the percentage of the total variables to fall in the hyperplane. Now, let us recognize at once that the percentage of variables to fall in such a hyperplane will depend on the nature of the population of variables and the nature of the factor. As a principle of good research design,

it is urged, in a later chapter (page 345), that one should attempt deliberately to insert into every battery a sufficiency of variables utterly unrelated to some of the main factors in which one is interested, in order to create good hyperplane fodder. The all too slowly dying misconception that one finds factor positions in rotation by putting reference vectors through clusters,[2] rather than by putting hyperplanes through disks or points is perhaps responsible for experimenters furnishing their batteries too lavishly with variables likely to be *high* in a factor, while omitting to make sure that there is a sufficiency of *low* variables to create a clear hyperplane by which the reference vector can be properly located. Actually both figure and ground in the picture must each receive proper attention.

Where the population of variables is deliberately chosen on the personality sphere principle or any similar principle of representing a wide universe of variables randomly, a number of studies (**21, 22, 27, 53, 102, 103, 125**) indicate that from about two-thirds to five-sixths of the variables may be expected to fall in the ± 0.10 hyperplane for any one factor, the former figure being obtained for ratings, which tend to have more width of reference, and the latter with objective test measures. Where there is deliberate concentration on some special area, but with a fair representation of marker variables from the chief known factors in other areas, for the sake of getting proper orientation, it seems that about a third or a half may be expected in the ± 0.10 hyperplane. Obviously where a greater number

[2] It is natural perhaps to assume that since a factor sometimes (but by no means always) corresponds to a cluster, it has been found by locating the cluster. This leads to the popular error that there are as many factors as there are clusters and that one can create factors by putting variables into a battery which one knows will stick together in a cluster.

Actually the factor or reference vector takes its position from the hyperplane, i.e., the things that are *un*correlated with the factor. A factor therefore can appear in the findings even if one does *not* have any cluster, intended markers, or strong representatives of it; though in these circumstances the loadings may be, inconspicuous, and the factor may not be easy to describe in test terms. From this fact, it is clear that to insert in a battery variables entirely unrelated to the factors in which one is interested is quite as important as to insert tests having the character of the factor. Ideally, *both* should be present to give both a clear hyperplane *and* a few very highly loaded variables to define the nature of the factor (as ground and figure). If the same tests can be good markers for one factor and good hyperplane stuff for the other so much the better, but at least it is important to have a sufficient percentage falling outside the factors one is primarily interested in, to provide this hyperplane stuff. Of course, directed hyperplanes cannot be formed by variables lacking loading on *any* factor.

of factors is involved there will generally be a greater number of variables in the hyperplane of each, and the above "two-thirds" figure applies to about a dozen factors.

It will be recognized that this standard is somewhat more exacting as to total hyperplane area than Thurstone's early criterion except in the particular case, with which Thurstone was mainly concerned, of most of the variables lying in a single personality area (in his case the area of ability). Our hyperplane area criterion also emphasizes the zeros per factor rather than the zeros per test, for individual tests may vary more widely in complexity than factors do in extent, providing variables are widely sampled. Naturally, in the whole matrix the percentage of zeros per test will *average* the same as the percentage per factor (when corrected by the ratio of the number of factors to the number of tests—which should be held constant in inter-matrix comparisons). The chief reasons for attending primarily to the factors are, first, the relatively trivial practical consideration that their fates are easier to keep track of in the rotation process and, secondly, the more basic theoretical argument that a factor is unlikely to be absolutely general and ubiquitous in its operation, whereas the argument that any single variable is unlikely to be complex is not so certain.

If one is working in a partly or wholly explored area, rotation is assisted by knowing that a certain factor is likely to have a well or a poorly represented hyperplane in the battery in question, i.e., by knowing how much hyperplane area to expect for a given factor, involving particular variables. For example, among personality factors, it is noticeable that general intelligence (B), dominance vs. submissiveness (E), and cyclothymia vs. schizothymia (A) affect so much of the personality sphere that they leave only just enough variables uninfluenced to create a hyperplane, whereas surgency vs. desurgency (F) and K factor consistently leave a large hyperplane. Thus in matrices of 35 and 36 variables, the numbers in the ± 0.10 range for these factors, when the best possible simple structure has been attained, have been, for three independent studies (see references in **30**) as follows: A: 15, 18, 17; B: 13, 22, 21; E: 14, 20, 21, while for F they are 18, 22, 24; and K: 24, 23, and 25. Similarly in the ± 0.05 hyperplane E averages 10 and A averages 12, while K averages 13.

In accepting guidance from the characteristic size of a factor's hyperplane, one must bear in mind that when a factor's variance

becomes small, more variables will naturally crowd into the hyperplane. In three studies with twelve factors, the first three in order of mean variance contribution (after rotation) have averaged 18 variables in the ± 0.10 hyperplane whereas the last three have averaged 22.5.

This raises in a new context the question of the loading limits that shall be accepted as demarcating the boundaries of the hyperplane. Our treatment of the question has so far been largely empirical, but to be complete must anticipate the theoretical discussion in Chapter 17 of the standard error of a factor loading and the effect of test unreliability upon a loading. The extent of "fuzziness" of the hyperplane through sampling error can thus readily be calculated to guide our choice of hyperplane width, but the experimental error revealed in reliability coefficients, being specific to each test, cannot be generalized to this purpose. Even when a theoretical estimate is used one can strongly recommend the practice of introducing *random variables* into any factorization, which function both as an empirical check on hyperplane width, for rotation purposes, and also as a check on other aspects of the factorization process. Random variables are purely artificial variables assigned with random scores in respect to the population. The correlations with the other variables are thus of chance magnitude, and though the random variables occasionally transcend the limits of the hyperplane (loadings of 0.3 and 0.4 having occurred in some of our studies) they generally indicate the range to be expected of variables which, apart from error, have zero factor loading. Naturally, to avoid distortion of the whole, and undue labor, only half-a-dozen are usually included in a 50 x 50 *r* matrix.

By whatever means the degree of fuzziness of the true hyperplane is estimated one must recognize that the number of variables falling within these agreed limits will also be affected by the mean size of all loadings in the particular factor concerned. Now there is no reason to believe that the true loadings will be other than normally distributed. Consequently, we can suppose that we have reached the hyperplane as we go from high to low loadings when the density of variables suddenly becomes higher than that expected from a normal distribution. In short the measure of how much is in the hyperplane might advantageously be determined by taking some ratio or difference between what is in, say, the ± 0.10 range and what would be expected to be

in that range having regard to the size (variance) of the factor and the distribution of loadings through the higher ranges.

The present writer has experimented (**21**) with some four purely empirical devices for an index of simple structure on any single factor and has found that rotation toward maximizing the ratio:

$$\frac{\text{Frequency of loadings from zero to half the mean loading}}{\text{Frequency of loadings from half the mean to the mean}}$$

to yield the best agreement with the positions most satisfactory on other grounds, but it is probable that some combination of this numerator with the absolute number in the ± 0.10 hyperplane (or some larger range if the error variance is at all appreciable) together with some regard for the probable hyperplane area of the particular factor in that particular battery (if prior or extraneous knowledge of its character exists) would be best.

It is this particularism of the structure, this need for combining several criteria, this refusal of nature to fit any obsessionally tidy mathematical scheme which makes rotation something of an art. It is these circumstances also which have been responsible for the defeat of attempts at purely analytical solutions of the simple structure problem. Nevertheless that same arduousness of the present methods which causes some researchers to fail in the pursuit of simple structure by single RV shift explorations has driven others to premature satisfaction with analytical devices. We must now glance, however, at those analytical or semianalytical devices which have some real practical use or theoretical promise, either alone or in combination with graphical methods.

ANALYTICAL AND SEMIANALYTICAL METHODS FOR ROTATION

The real antithesis in rotation methods, incidentally, lies between analytical methods and trial-and-error methods, but it so frequently happens that the latter are graphical that this graphical-nongraphical distinction often appears conterminous with it. The ideal analytical method would be one which would arrive directly at the transformation matrix required to give simple structure through the solution of some equations describing the correlation configuration. It would determine the position of existing hyperplanes, despite their being blurred by error, by the device of finding some expression to give a best fit when minimized. For example, we should be most likely to

choose a least squares fit, i.e., to write an expression for the sum of the squares of the projections of the supposed hyperplane variables upon the reference vector in question and to set up an equation which would show us when this expression is a minimum.

Unfortunately, by any of the simpler, known mathematical usages, we should need to know beforehand *which particular set* of variables is to be taken into the expression in order to find the position that would minimize the projections. Granted that we know the variables, rotation could then be carried out immediately without trial and error by analytical means. For example, Horst (**75**) has suggested that we consider a reference vector to be settled on its hyperplane when the expression

$$\frac{\text{Sum of squares of projections of highly loaded variables}}{\text{Sum of squares of low (hyperplane) loaded variables}}$$

reaches a maximum. But this requires that we somewhat arbitrarily decide when a loading is high enough to be high and so on. It also requires, as does any existing analytical method, that we first know which variables are going into the hyperplane, and that is precisely what we want to find out! Conceivably, if axes and hyperplanes could be kept orthogonal, a mathematician could invent an expression to handle simultaneously the total of projections (for all factors) that are of *limited* amount (say within ± 0.10). These would constitute the hyperplanes, and we should determine analytically the rotation which would bring this to a minimum. When the complication is added that the hyperplanes to be found are actually oblique and of an unknown degree of obliqueness, and that different hyperplanes contribute different amounts to the total, the problem defies—or at least has so far defied—solution.

However, there are advantages in using the solution of equations instead of graphs even if the method in fact is only semianalytical, requiring indeed a prior choice, of a mathematically arbitrary kind, of the variables to be included in the tentative hyperplane. Thurstone (**125**, Chapter 17) has designed and gathered experience with two such semianalytical methods, with which, however, it is advantageous to get supplementary aid by guidance from graphs. They are essentially the same device, one using an unweighted measure of the divergence of tests from the hyperplane, and the other a weighted measure. They follow the line suggested above of minimizing the

projection of certain tests upon a reference vector but they do not overcome the difficulty that the set of tests has to be arbitrarily chosen. Consequently, they do not proceed by a single computation to the best simple structure position, but converge by successive approximations, groping to the right set of tests. For as the simple structure is approached, it usually becomes evident that the initial arbitrarily chosen group for the hyperplane is not entirely correct and must be changed. In this respect the methods do not live up to the requirement of our pure analytical solution, but they work pretty well, and in the hands of a highly experienced worker are often better than the graphical method, because they are significantly quicker.

The steps will be described only in briefest outline here, the student being referred to Thurstone for the full description and justification (**125**, Chapter 17). The methods begin by rotating by a single jump, as described on page 204 above, from the unrotated matrix to a set of oblique trial reference vectors placed along certain chosen test vectors. The duration of the subsequent process and its claim to be quicker than the graphical process depends a great deal on the choice made for these initial trial vectors. One then makes from the calculated V_1 matrix a table of *distributions* of loadings on each *RV* and chooses a set of variables that form a mode on each, i.e., that cluster with the highest frequency in regard to magnitude of loading. Thus on one it might be that most variables fall in the loading range $0.2-0.3$, and this particular set of variables would then be chosen as the group to form the hyperplane to have their loadings minimized. Usually, if inspection of the column in the unrotated matrix shows that there is some possibility of a choice between rival high and low modes, one takes the lower bunch of projections; a higher set would indicate that too large a swing of the *RV* has to be carried out to minimize them. Alternatively this choice of a group can be carried out by drawing the usual graph and looking for a nebula of points.

But in this case there is a difference from the usual inspection of a graph. Normally the nebula must run through the origin, and if that happens here we can use the calculation with which we are already familiar (a shift by λ matrix calculations) to bring the *RV* to the right position. But if it does not—and on the first, V_0, matrix a nice line seldom presents itself—the present method has the advantage

of being able to find the *RV* direction cosines which will put a hyperplane as nearly as possible through *any* set of points, regardless of whether they form a line *through the origin*. After such a shift (the calculation for which may be read in Thurstone **125**), a graph is almost essential, for one now needs to see any reclustering of points that may be occurring to suggest what might be added to the hyperplane group to be pulled in by the next shift. Usually three or four shifts suffice to gain simple structure, as compared to say, nine to twelve on the unmodified single-view graphical method. But this saving in moves is bought at the cost of more calculation per move and of the time of a more highly skilled worker than is required for the other method.

The weighted method is similar except that the sum of projections to be minimized is arrived at by weighting some tests more than others. The tests that are positive and near the hyperplane are to be made more negative, those that are negative and near the hyperplane are to be made more positive, and those that are remote from the hyperplane and which are unlikely to move into the hyperplane at all are to be left as far as possible unchanged. Perhaps to an even greater extent than in the unweighted method its success depends upon starting from a position relatively near to true simple structure, for one explicitly sets out to pull into the hyperplane the tests that happen to be near the trial hyperplane. Of course, if the whole succession of approximations lead nowhere, one can start again from some other position. But it is not so easy to decide here as when guided by the plateau record (history of the hyperplane) of the simple graphical method, when, in fact, one *has* gotten nowhere, i.e., to decide whether the hyperplane attained is really an unsatisfactory one. Only the absolute number of items can now be used as a guide. Consequently these methods are probably best used either when one has good reason for confidence in the choice of initial trial reference vectors or when preliminary graphical rotation has already begun to show some condensation of nebulae of test points into prospective hyperplanes.

METHOD OF EXTENDED VECTORS

Another method which, though not analytical in nature, aims to discover the hyperplane without the extended groping of the simple graphical method is that developed by Thurstone (**125**), Harris (**67**),

and others in various uses of extended vectors. To illustrate the general nature of this approach, we may take three oblique factors—F_1, F_2, and F_3—that are somewhat positively correlated, as in Diagram 24. It will be remembered, incidentally, that the factors are the intersection lines of the hyperplanes, not the reference vectors perpendicular to the hyperplanes. Next we take a screen (a plane) and hold it some little way off the origin—say, at unit length of one of the factors and at right angles to it. (In Fig. 1, Diagram 24, the screen cannot clearly be shown at unit length, but it is supposed to touch the sphere at factor axis F_3 to which it is supposed to be perpendicular.)

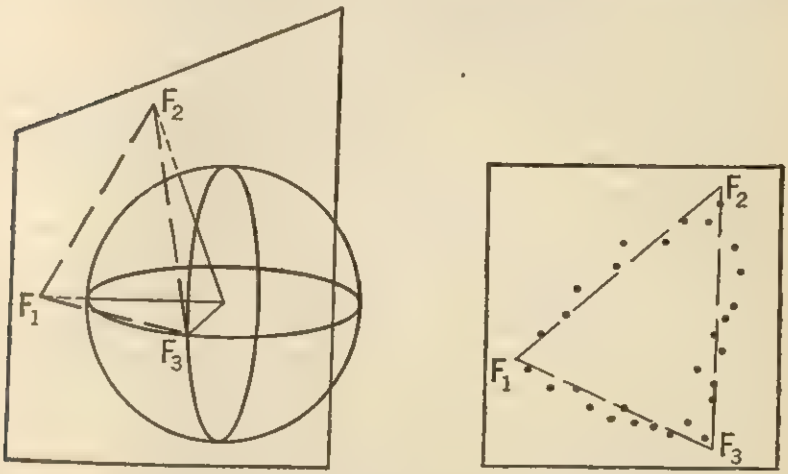


Fig. 1 Fig. 2

DIAGRAM 24. Hyperplanes by the Method of Extended Vectors.

The extensions of the test points which form the three (hyper) planes would cut it in approximately straight lines as shown in more detail in Fig. 2, Diagram 24.

The diagram corresponding to that on this screen can be produced by relatively little calculation from the unrotated factor matrix, and inspection of it should show these lines of denser points relatively easily, after which the positions of the corners (intersection of planes), i.e., the factors, can be read off with respect to the unrotated factors. With more than three dimensions the problem becomes a little more complicated. Harris (**67**) suggests that only four planes can be located on a single diagram, so that with more dimensions several overlapping

views must be taken. Various difficulties remain to be worked out, and at present these methods are probably not as practicable for routine computers as are others mentioned here.

PRINCIPLE OF PARALLEL PROPORTIONAL PROFILES

An attempt at a purely analytical solution which has theoretical promise but which has not yet been cleared of practical difficulties is that known as the method of *parallel proportional profiles* (**18**). It attempts to put into a new operational form the basic principles stated earlier—that parsimony should be sought simultaneously with respect to several matrices rather than just one. It thus assumes that if a functional unity is real, it will appear as a similar factor, i.e., a factor with a similar loading pattern, in several different factorizations. The design thus calls for starting with at least *two* distinct experiments, each using the same or overlapping variables. Upon rotation of the findings there should be just *one* position in which the loading pattern in one will correspond to the loading pattern in the other. It can be shown that when this is true of one position no other positions can be found for which this is also true (**18**).

The problem in practice is much like that of finding the key to a combination lock. Two dials can move into any position and the solution requires the simultaneous discovery of two uniquely related positions. It might presumably be solved by trial-and-error search, perhaps of a systematic kind in which one rotates the first matrix steadily through a series of positions at each of which the second is moved through all possible positions, until the pair clicks into a solution. But this would be an almost infinitely long task and, as we shall see, an analytical solution is possible. The term "corresponds" needs amplification. In the very special case in which the variance of all factors is *exactly* the same in both matrices no solution is possible, for it can be shown that any position of one could then have a corresponding pattern—in fact, an identical loading pattern—in the other, i.e., a set of similar loading profiles with respect to all factors. It is necessary that the two experiments have the same variables and yield the same factors *but that the variance of each factor in one experiment shall be different*, through accidents of sampling or through deliberate manipulation of experimental conditions, *from that of the corresponding factor in the second.*

For example, the variance, i.e., the mean square of all the significant

loadings, of a set of intelligence measurements would be lower in a well-selected college population than in the general population. Thus presumably all the loadings in the pattern would be *equally* or *proportionally* reduced for a given factor in the one experimental situation; in this case in the sampling situation of having a more restricted range. The loading profiles would then be parallel or proportional[3] (or proportional at parallel points) in the two rotated matrices when the true position is found.

It is possible, by a rather complicated solution of simultaneous equations, to arrive analytically at the rotations from any two experiments that will give proportional profiles (**34**). Unfortunately it remains to be demonstrated by an adequate practical example (see **34**) that the chance errors and the complication of the proportionality of profiles, beyond that posited in the simplest hypothesis stated above, will permit this analytical solution to work.

OTHER PRINCIPLES BEYOND SIMPLE STRUCTURE

In the present theoretical discussion, dealing with a rotation method which goes beyond the goals of simple structure, to which the practical methods discussed above have always previously been subordinated, it behooves us to take stock of a variety of other methods which also work (or might work, for in general, they also fail) on different assumptions.

Some half dozen have been tried, mostly before the time of simple structure and with none too explicit assumptions. A common device among the unsophisticated is to rotate for psychological meaning, i.e., move a factor to a position where the high loadings agree with some preconception of the experimenter. Almost *any* preconception can be confirmed in this way, for rotation is a flexible tool. This approach merely perpetuates erroneous speculations inherited from the crude infancy of psychology, e.g., the belief in an extraversion factor.

A second approach may be called the *principle of orthogonal additions.* Here one has to start with at least one factor that is admittedly known. One endeavors then to add new tests to the battery that will introduce *only one new common factor.* Since even oblique factors are

[3] More research is needed to decide the exact nature of this proportional or parallel relationship. Presumably a fairly complicated relationship would constitute matching here; but for the present, to simplify the solution, we shall assume the loadings in one pattern would hold a fixed ratio to the loadings in the factor matched.

generally *approximately* orthogonal, one can locate this new factor reasonably well by setting up a reference vector orthogonal to those already known. This was done in the early work of Webb (**136**) and of Garnett (**57**), where Spearman's g constituted the known factor and the new dimensions they found, and which would now be called emotional maturity or C, and surgency or F, were placed at right angles to the factor.

A third device, that of putting axes through the center of clusters is one which, if the general principles of simple structure are sound, is basically misleading. It dies hard because in more than a chance fraction of cases it happens to give the same solution as simple structure. For a cluster may be an overlap of two or more factors, but it *may* also correspond to a single factor. On statistical grounds alone it is a faulty method, for by arbitrarily putting in a battery a set of very closely related tests the experimenter can make clusters wherever he likes.

A fourth approach—rotation to agree with past factor analyses—is very similar to rotation for psychological meaning, i.e., it aims to agree with past clinical syndromes or preconceived psychological entities. Psychological meaning, in the narrower sense of agreement with past *factor analyses*, at least when the past factor analyses have been based on some better principle, *can* be useful, however, in special circumstances. This utility is shown most in applied studies, which are not attempting to break new ground in basic research, but the method obviously cannot create new knowledge, except by revealing the factor composition of a few new variables not used in the earlier batteries.

Some claim can be made for a fifth method, that of rotating to obtain certain general characteristics—other than simple structure—which might be expected of psychological factors. It can operate upon a single matrix and within the limits of a single experiment. For example, in a certain situation one might expect all factors to have approximately equal variance rather than diversified variance, or one might expect plateau loadings (**18**), i.e., a pattern in which some variables are highly and equally loaded and all others negligibly so, rather than a normally distributed set of loading values. Such criteria include the factor constellation forms discussed in Chapter 9, where some investigators believe in the likelihood of certain general forms being more real. Nothing in this area of a sufficiently exact and par-

ticular nature to guide rotations has yet gained rational, empirical, or professional support in this field.

A last principle of rotation is not to rotate at all! Such an economical procedure naturally has its devotees. The most systematic defense of the procedure has been made by Burt, followed by Eysenck. The only substantial claim yet put forward, apart from economy, is that it can be empirically shown to give factor invariance just as well as simple structure or any other adequate method. On grounds of experience this cannot at the moment be adequately proved or disproved, since insufficient repeat factorizations of the same variables under adequate experimental and sampling conditions exist today.[4] Those which *do* exist, in the writer's opinion, show more indubitable invariance for simple structure than for nonrotation.

However, on theoretical grounds, the use of the unrotated matrix, involving a first general factor, with no zero loadings and usually all positive loadings, does not in the first place agree with the general psychological expectation that all factors will be of the same general species of loading pattern and the same general order of mean variance. Further—and this is more surely fatal to the notion that such factors can be invariant—one has to recognize the statistical fact that, except for one unique situation (when the hierarchy of correlations is not upset), the addition of new variables to the battery alters the centroid of those preëxisting. Their loadings in the first (and in all subsequent) factors vary according to the company they keep. This makes the use of unrotated factors psychologically and statistically meaningless.

The last comment applies, of course, both to the unrotated centroid and the principal components solution. These differ, as is brought out most clearly by Peters (**153**) and Burt (**11**), only in that the former makes the *sum* of the factor loadings a maximum, whereas the latter makes the *sum of the squares* a maximum. In both the nature of the first factor is influenced at once by adding a new test to the battery, and in both the succession of bipolar factors, equally positive and negative, which follows, is likewise affected. Neither

[4] There are only two regions which supply examples at all: the primary abilities, as analyzed independently by Thurstone (**125**), Bechtold (**4**), Carroll, Goodman, Meili, and Rimoldi (**103**); and the personality sphere in terms of ratings (see references in **22**) and to a less definite extent in questionnaire studies where Guilford's analyses are borne out by other studies (**30**).

method thus gives factors of invariant scientific meaning unless rotated to simple structure.

Among the few *psychologists* who have not yet adopted movement from the unrotated to a simple structure position, e.g., Burt, Eysenck, Stephenson, Reyburn and Taylor, various hybrids of the above alternatives have appeared. But only in one case—Eysenck's criterion analysis (**49**)—has a principle appeared requiring additional description in our series. Since the term *criterion analysis* has been used in a more general and appropriate sense by other factor analysts, as described in Chapter 20, the specific process involved here is better called *criterion rotation* or *criterion oriented rotation*. It consists in including in the factorized population two distinct homogeneous groups, e.g., normals and neurotics, advanced and retarded school children, presumed to differ in the factor presumed to account for most variance in the test battery. Subjects are given scores according to the group to which they belong, e.g., normals 0, neurotics 1, and this criterion variable is biserially correlated in with the other, continuous variables. The factor of presumed "neuroticism" in the resulting factor matrix would then receive final rotation to be collinear with this neuroticism variable.

The logical objection to regarding this as an independent, sufficient method is that the only proof one would have that the two subgroups differ only in the factor concerned would derive from previous measurement of them on the factor. Most group means differ in respect to more than one factor, e.g., advanced and retarded children do not differ systematically only in intelligence, as is recognized, for example, in using the discriminant function. Consequently, a factor so obtained is likely to be anything but pure. Criterion rotation belongs in the same general category as peripheral validation (**22, 30**). It gives additional meaning to a known factor and may help in *naming* it to the satisfaction of applied psychologists. For example, if a certain factor correlates so highly with the criterion of neuroticism that no other factor could account for so much of the variance, the proof that the simple structure rotation thus aligns itself with the criterion variable adds justification to our calling the factor one of general neuroticism. As Eysenck points out this permits one to test a hypothesis, but the "hypothetico-deductive method" is no *more* involved here than in rotation for simple structure. In the latter one experiment suggests

a hypothesis about the nature of a factor and another, entered perhaps with additional variables, enables one to test it by seeing if simple structure yields the hypothesized loading pattern.

REVIEW OF PRINCIPLES OF FACTOR RESOLUTION

Of the eight methods surveyed, only simple structure and parallel proportional profiles (which may be regarded as a kind of simultaneous simple structure) are theoretically sound for general application. Orthogonal additions, page 247, has a legitimate but limited role. In theory, proportional profiles is the ideal method, for simple structure has three drawbacks:

1. The definition and clarity of the hyperplane is intrinsically poor in some factors. By the very nature of social, biological, and psychological influences we sometimes encounter a factor which is so general and so potent as to affect practically all variables.

Again, the tendency among research workers to be consciously or unconsciously preoccupied with a limited area of events often causes them, even when fully aware of the necessity for introducing hyperplane stuff, to work with batteries of variables in which some influence or influences are common to the whole matrix. At least one truly general factor, devoid of a hyperplane, then confuses attempts at rotation. The second weakness is probably more common than the first. The first yields a vague hyperplane of limited area; the second yields none at all.

2. Alternative simple structures are indubitably sometimes found, and in the absence of any generally acceptable criterion we cannot decide, from the given matrix alone and without extraneous evidence, which is preferable.

3. No analytical method of shifting directly to simple structure seems likely to be found in view of the "relativity" of simple structure, but various analytical computing aids can be used for certain steps in the rotation process so that whatever intellectual attraction may exist in the game of hide and seek for simple structure shall not be completely offset by the laboriousness of the graphical rotations.

However, the proportional profiles principle is at present not in a practicable state, and simple structure remains the only theoretically acceptable and practically well proven method available. With ancillary aids in the special cases where it limps, it is fortunately able to meet the demands of practically all experimental designs and situations.

Questions and Exercises

1. Discuss two or three criteria for deciding when a proper hyperplane for a factor has been obtained. What are the characteristics of a good hyperplane?
2. What considerations determine the limits between which projections are to be considered to lie within the hyperplane, and what expression can be used to allow for the magnitude of mean loading of all tests in the factor?
3. In a group of tests measuring known traits or conditions, how can one predict in advance whether certain factors will have large or trivial hyperplane areas? What is the advantage of including in the test battery tests known to be both high and low in these factors, rather than allowing the factors to identify themselves without being so guided? How and why is hyperplane stuff introduced into a battery of variables?
4. Describe briefly the aims and methods of attack upon rotation by analytically minimizing certain groups of loadings, or projections, on each factor. What are the advantages and disadvantages of such methods as compared with graphical trial and error rotation?
5. Discuss the possibility of purely analytical methods, and describe in some detail the aims and present practicability of the method of parallel, proportional profiles.
6. What is the general method of the principle of orthogonal additions?
7. Summarize some other methods of rotation not mentioned in questions 1 through 5, and point out advantages.
8. Describe as many historical instances as you can of factor invariance, i.e., the rediscovery of the same factor pattern in different researches, and discuss to what extent these support the validity of the simple structure criterion vis-a-vis other criteria of factor resolution.

CHAPTER 15

The Basic Art of Rotation by Graphs

In the last chapter we have made a comparative survey of the chief methods and criteria available for the process of factor resolution by rotation. Our purpose now is to concentrate upon the practical skills required in the basic method of rotation of single *RV*'s by graphs; for it is this method which is most used in the majority of laboratories and which is needed as an adjunct even when one uses technically more specialized, analytical methods.

THE ROLE OF SPATIAL REPRESENTATION

We shall proceed, therefore, to describe the fullest possible developments, in accuracy and the use of facilitating short cuts, of the graphical method, the essential aim of which is to keep good visual control. Even for the somewhat unusual mind which prefers algebraic to geometrical proofs the visual presentation in this case permits a quicker overall view of what is happening. Throughout factor analysis we have to recognize that we are dealing with an art or craft and that its proper use depends on wise judgments based on experience. But it is in this process of *rotation* that skills of an artistic nature, not communicable by mechanical instruction alone, become paramount. This will cease to be true only when a purely analytical solution is invented.

A resourceful reader may wonder at this point why, if we are to handle the whole matter by diagrams, we do not make a three-dimensional model as an instrumental aid offering more direct solutions than the two-dimensional graph. This has been tried by the present writer's laboratory (**91**), using a system of peg holes for recording the loadings in two dimensions and various positions of balls on rods for the third dimension. Short of suspending the whole on gimbals, as in a mariners' compass or climbing on the walls and ceiling to view a fixed model, it proves not possible, however, readily to get the eye in a position to find the hyperplanes. Thurstone has used the alternative possibility of extending the test vectors and putting chalk marks on the surface of a solid sphere, where they cut. To obtain these positions

requires calculation from the given loadings, and unless the sphere is also transparent it is not easy to see the hyperplane of points. In short this approach sounds much simpler than it turns out to be. And since few problems reduce actually to three factors and the task of reintegrating results with the other factors is rather more awkward with three than two factor views, there is nothing greatly to recommend it. Most of those who have tried it have returned to the method of taking only two factors at a time and looking at the sectional views thus obtained on graph paper.

Using graphical methods does not mean—except for the roughest work—that we dispense with calculation, as the student already realizes. The graphic views are to guide the calculation process and render its results visible, but the shifts are actually fixed by calculation. Thus whether we use sectional views[1] or some more analytical method, the process of matrix multiplication plays a recurrent and sometimes major role in the computations. For this reason the student who is going to spend much time in factor analysis does well to familiarize himself with the principles of matrix algebra, either through the mathematical introduction to Thurstone (**125**), or Holzinger and Harman (**71**), or through one of the briefer mathematical treatises, e.g., (**54**). For the student who does not proceed to the point where it is helpful to have all the tricks of the trade in reserve it may suffice to understand matrix multiplication as it has so far been encountered here—namely as a means of getting through a lot of individual multiplications in an orderly fashion, where the individual multiplications make sense in terms of ordinary trigonometry. The matrix rules for postmultiplication (column of second by row of first); for setting out a transpose (row becomes column) and for calculating an inverse are the only ones the user of the processes of rotation needs to know thoroughly.

FACILITATING ROUTINE ROTATION COMPUTATION

Before turning to the arts and skills of the graphical process as such, let us summarize and describe the practical devices for the accompanying calculation processes themselves. In accordance with

[1] We shall use the expression *rotation by sectional views* in conformity with other writers to mean the taking of two-dimensional graphical views followed by rotation of one or both axes. The expression single-plane rotation which might seem to mean this has been specialized by Thurstone to apply to the semi-analytical method of shifting a single reference vector and its hyperplane, which is described here in Chapter 16.

the description given in Chapter 13, the basic computations in rotation to an oblique position may be set out as in the following table for a factor matrix with n variables and k factors carried through m rotations on all reference vectors,[2]

TABLE 33.

Unrotated factor matrix V_0

	F_1	F_2	F_3	$..F_k$
v_1				
v_2				
v_3				
.				
.				
.				
.				
.				
.				
.				
.				
v_n				

×

Transformation matrix λ_m

(Unrotated factors) \ (Rotated factors)	F_1^m	F_2^m	$..F_k^m$
F_1			
F_2			
.			
.			
.			
.			
.			
F_k			

=

First rotated factor matrix V_m

	F_1^m	F_2^m	F_3^m	$..F_k^m$
v_1				
v_2				
v_3				
.				
.				
.				
.				
.				
.				
.				
.				
v_n				

while the direction cosines of the new reference vectors one to another are calculated by obtaining the C matrix as follows:

TABLE 34.

λ'

	F_1	F_2	F_3	$\ldots$	F_k
F_1^m					
F_2^m					
.					
.					
.					
.					
.					
.					
F_k^m					

×

λ

	F_1^m	F_2^m	$\ldots$	F_k^m
F_1				
F_2				
.				
.				
.				
.				
.				
.				
F_k				

=

C or $\lambda'\lambda$

	F_1^m	F_2^m	$\ldots$	F_k^m
F_1^m				
F_2^m				
.				
.				
.				
.				
.				
.				
F_k^m				

[2] It is, of course, realized that the use of F_1' for the first factor in the rotated matrix (or F_1^n for any later rotation) does not indicate any special relation of this factor (other than an historical one) to F_1 in the unrotated, except that, if we start other than with trial vectors, F_1' was *originally* close to F_1.

The conversion to angles among factors, in the C_F matrix, and loadings upon factors, in the F_m matrix, is summarized similarly later on, not being part of the rotation as such.

Since in the actual process of multiplication computers find it convenient to carry a row of multipliers on a strip of paper as they multiply row after row of the unrotated matrix, it saves a little time and some chance of error *if the λ matrix is set out each time in the transposed form*, so that the *rows* of multipliers are taken directly from it. This keeping of what are really columns in row form is also more convenient in actual computation when calculating the changes in direction cosines occasioned by each shift, as shown in the standard *rotation computing form* set out in Table 35.

In this table the λ matrix as it exists up to this set of drawings is written in the upper matrix, F' cosines being written horizontally, in rows, as just stated (up to 14 factors). The number nth written before Rotation at the top is that of the rotation just made which has yielded the λ matrix written in the top frame, the $C = \lambda'\lambda$ matrix written in the bottom frame, and the drawings which are just being studied, and which have these C angles written on them. The next step is now to fill in the required shifts in the center frame. (Incidentally, starting with trial vectors will automatically start this record form in the correct phase, but starting with a shift directly from the unrotated matrix would give transformations for shifts which yield the λ and C of *the same record form* instead of the following one, unless one took care to interpolate an extra record form.) These transformations for shifts are usually written in this form:

$$F_3^{n+1} = F_3^n + 0.2F_7^n - 0.4F_9^n \tag{26}$$

n being of course a superscript, not a power, and the 0.2 and 0.4 representing concrete instances of tangents read off from two graphs on which a shift has been seen for F_3, namely, on graphs F_3 by F_7 and F_3 by F_9. The addition and normalization of the three rows of cosines corresponding to the three terms on the right is made, as has been described, on scrap paper retained only until the next set of drawings are made, and the results are entered in the top frame of the next, the $(n+1)^{\text{th}}$, rotation record form.

Machine aids in matrix multiplication are yet in their infancy, but when developed will greatly reduce the heavy labor now involved. A modification of the I.B.M. machine by Tucker (**129**) enables a

TABLE 35. Computing Record Form for Rotations

*n*th Rotation transformation matrix

		1	2	3	4	5	6	7	8	9	10	11	12	13	14
F^0															
F^n	1														
	2														
	3														
	4														
	5														
	6														
λ_n	7														
	8														
	9														
	10														
	11														
	12														
	13														
	14														

Ensuing transformations for shifts

F_1		F_8	
F_2		F_9	
F_3		F_{10}	
F_4		F_{11}	
F_5		F_{12}	
F_6		F_{13}	
F_7		F_{14}	

		1	2	3	4	5	6	7	8	9	10	11	12	13	14
	1														
	2														
	3														
	4														
	5														
	6														
C_n	7														
	8														
	9														
	10														
	11														
	12														
	13														
	14														

whole row by column multiplication to be carried out at one operation (by electrical summation of the individual multiplications) but such machines are not generally available. As pointed out in Chapter 19, the new electronic calculators promise to do the *whole matrix* multiplication at one operation, though with matrices of more than, say, 20 variables and 6 factors it may be necessary to work in two steps, because of the limit to the number of digits the machine can remember.

STEPS AND DEVICES IN GRAPHICAL ROTATION

Certain aids which apply mainly to the sectional-view, single-reference vector method of rotation but also to some others will now be briefly described. First we may note the device of shifting the *RV* on two or more other *RV*'s at once as implied in the transformation entered on the record form above. In a previous chapter it was pointed out that if a reference vector F_1 is shifted through θ degrees *toward* reference vector F_2, its new direction cosines are obtained by normalizing the inner sum of F_1's old direction cosines and the direction cosines of F_2 multiplied by tan θ. Now if one finds suitable shifts for F_1 on three drawings, say, with F_2, F_4, F_5 (the tangents of shift equaling .20, −.15, and .23 respectively), the shifts can be added giving a calculation of the following form (in a six-factor problem):

TABLE 36. Combination of Several Shifts

	F_1^0	F_2^0	F_3^0	F_4^0	F_5^0	F_6^0
F_1 =						
$+0.20F_2$ =						
$-0.15F_4$ =						
$+0.23F_5$ =						
Total equals.......................................which normalized						
becomes.......................................						

However, in combining shifts in this way, especially if the shifts are made from positions already oblique, there is danger of missing the intended position and especially of overshooting it. (For the influence of the original position—reference vector F_1—in the above summation is reduced by the cumulative influences of several new additions.) When the identifying numbers of the individual variables are written alongside the points in the drawings (and this need not be incompatible with blind rotation, as advocated, page 409), it is possible to see to

what extent the points being approached by the proposed new hyperplane in one drawing are the same as those being approached by the new hyperplane in the other. If they are absolutely identical, there is no point in making both shifts (even if one did so as two half-angle shifts which would not overshoot the mark) unless it is necessary to improve the angle between the reference vectors by such shifts in *both* cases, e.g., to reduce oblique angles that have become too big.

Incidentally, although there has already been ample discussion of the pros and cons of keeping the whole system roughly orthogonal and avoiding extreme angles, a word may appropriately be added in this

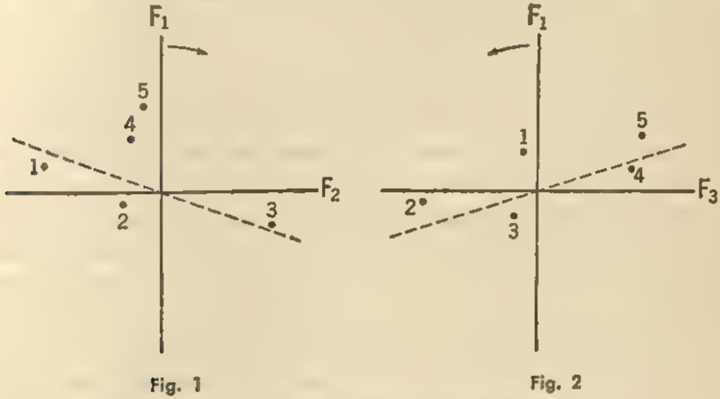


DIAGRAM 25. Advantageously Combined Shifts.

more detailed discussion of the single section method on the necessity for constantly keeping an eye on the C_n matrix. In the first place, when only a fraction of the possible drawings are made, that fraction should generally include those pairs with bad angles—say, above ± 0.4. Second, the C matrix is the simplest place to keep a record of what drawings have been made by encircling the angles corresponding to the drawn graphs after entering them on each graph. Lastly, when in doubt as to which of two hyperplanes in a drawing one shall shift—especially when the shift is partly dictated by a desire to reduce the angle—it is generally best, other things being equal, to shift that which is shown by the C matrix also to have poor angles with *several* other factors, for this means that it is probably still quite unsettled.

As to combined shifts, the most ideal situation is as shown in the two figures of Diagram 25.

COMBINED SHIFTS

The shift of reference vector F_1 toward F_2 (Fig. 1) will bring variables 1 and 3 into the hyperplane. These will not be moved out again by the shift proposed in Fig. 2, because 1 and 3 are near the origin. The shift of F_1 away from F_3 will, in turn, bring 4 and 5 into the hyperplane without these being lost by the shift on F_1. Another important relation to observe is that represented by variable 2. The desirable shift in Fig. 1 makes its loading, already slightly negative, still more negative; but this is corrected by the shift in Fig. 2 which makes it more strongly positive, so that it finishes up in the hyperplane.

If one has to spend too much time in comparing the fate of individual variables in different drawings in this way, more is lost than gained by such recourse to combined shifting, even if successful. But if the hyperplane being approached is really the ultimate hyperplane, the relations that are likely frequently to appear are precisely of the kind indicated here, and with practice they can be recognized at a glance. Combined shifts thus have their maximum value when the two or more moves are definitely supplementary. They are safer to make in the later stages of a rotation when the final, required positions are fairly close. On the other hand, they *can* do much good also in the early stages, where an *RV* clearly needs to take up a new position at once relative to several *RV*'s, but the process is then more risky and it is best consistently to underestimate the shifts in such circumstances. For combined *large* shifts are apt to blow up and lose completely whatever semblance of a hyperplane had been gained up to that point.

EMPHASIS ON VARIOUS DEVICES

In such decisions between alternative rotation practices, just as in the similar problem of the factor extraction processes, we have to weigh alternatives which ought properly to be examined by a detailed cost accounting of associated time expenditures. Thus this issue of combined shifts is tied up with the question of whether it is economical to make *all* possible drawings at every round of shift, e.g., 66 graphs in the case of 12 factors. The alternative is to stop making drawings containing a given factor as soon as a good shift for that factor is seen. The latter practice is probably less satisfactory, for in searching for good shifts for recalcitrant factors one usually has to make more

graphs on earlier factors already placed and these placings may then appear premature. With practice, moreover, on standard graph paper and a standard system (e.g., always putting the factor of lower index number vertically, and the angle in the top left corner) most people can reduce the total time for graph drawing to a fraction of that for computing, at least if unnumbered points are used. Numbered points can be added in those few cases where combined shifts are calculated. However, although making combined shifts on two indicated, "unnumbered" hyperplanes, which may or may not be views of the same real hyperplane, is a gamble, it is one generally worth while. As indicated above, it is more of a gamble in the early stages of rotation. In perhaps a third of such cases the two proposed shifts are actually aiming at different hyperplanes (corresponding to two different eventual factors). Then one falls between two stools and has to return to one or the other, losing the work of one or more shifts.

When all drawings are made, one is certain at any rate of having found the best *available* shift and only in some cases will one have the *embarras de richesse* of having shifts so equally good that they must, if possible, be combined. In this combination it is advisable, in addition to observing the principle already stated, (1) to combine on not more than two or three single section moves, (2) to combine three or more only in the latter half of the rotation history and with shifts that are small (preferably 0.1 or 0.2 but occasionally up to 0.4), (3) to underestimate slightly the shift on each, as indicated before, and (4) to favor inclusion of shifts which improve the angles between *RV*'s. It should be noted apropos of unexpected results from a rotation that even a single shift will frequently not yield *exactly* the result one would expect from the drawing, due to the fact mentioned before that our drawings are approximations putting oblique loadings on orthogonal graphs. For this reason, if the correlations between reference vectors rise above about 0.6, it is preferable to make true oblique graphs. These will be made by putting the *RV*'s at the angles found in the *C* matrix and using polar graph paper or ordinary graph paper with a strip of it tacked along one edge of a set square which is then run up and down along the axis that is drawn obliquely to the set of the graph paper, as mentioned earlier. For the projections as given on the *V* matrix are, for the *RV*'s, *perpendicular* projections on the oblique axes. (A glance at Fig. 2 in Diagram 26 may clarify this point.)

There are conditions, however, when all possible interfactor graphs are not made. For example, some three or four factors may achieve definite hyperplanes early and have no further need for rotation. These should nevertheless be used in further drawings from time to time because of slight polishing improvements that will crop up in them and because they may give indications of the true position of some of the less stable factors—a position likely to be roughly at right angles to them. If not all factors are drawn, it is good to keep an eye on the history of the hyperplane table in order to circulate the factors in all possible combinations, for too much pairing of particular factors cannot bring in new variance and radical improvement. Granted attention to this principle one can advantageously follow a second principle, due to Saunders (**108**), of pairing at each round those factors which tend to have the same variables in their hyperplanes. For this means they share a clustering of points about the origin so that each is free to rotate to pick up variables lying out at the periphery without losing the majority of members of the existing hyperplanes.[3]

ECONOMIES IN THE ROTATION PROCESS

When factors are numerous, the task of finding a structure is apt to be complex and to require much patience, especially in the early groping stages. One approaches a hyperplane and loses it again, or waits a long time for even a single hyperplane to crystallize, especially when the initial trial vectors happen to be badly chosen and movement is delayed by the generally necessary practice of trying to keep the whole system roughly orthogonal. But the labor, if not the complexity, of rotation can also be increased by sheer numbers of variables as much as by the numbers of factors. However, with large matrices, e.g., of fifty to a hundred variables, especially if there is reason to believe that all hyperplanes are tolerably well represented among the variables, it is quite practicable to carry out the rotation with, say, a half or a third of the total variables. These can be chosen deliberately to give the greatest variety of meaning and representation or by mechanically taking every other variable, as in split-half test divisions. It has been found in such circumstances that

[3] If H_1 is the number in one hyperplane and H_2, is the number in the other, the percentage of all variables common to both would be expected on chance to be $H_1 \cdot H_2 \cdot 100/n^2$ where n is the number of variables. This figure can be taken as a criterion for deciding when the number is actually in excess of chance.

the general position reached after three or four rotations of the whole system using the graphs for the even variables is well sustained when one substitutes the odd variables. Indeed one can safely proceed almost to the final rotation depending on calculations and drawings made for only half the variables, alternating the odd and even populations either every rotation, or perhaps with greater ease, every three or four rotations.

Another possible shortening of the rotation process arises from the fact that the errors of position of variables in the intermediate graphs will not, if slight, make any systematic difference to the rotations chosen and will not accumulate to the final solution—since one always refers back to the V_0 matrix. It *is* possible to arrive at the projections resulting from a shift of axes *without* calculating the new λ or C matrices, either through proceeding by a direct calculation from V_n *to* V_{n+1} or by a graphical transcription process. The latter is a real saving and will be described in detail, but the former merits only a brief indication, because it turns out to be as long as the usual process and has real use in only a few special circumstances.

In what we may call *the direct factor matrix transformation*, if we have shifted, say, F_6 to the extent of θ degrees toward F_{11}, we should normalize the two terms 1 and tan θ, which might result in, say, 0.8 and 0.6; and multiply in the V_n matrix the F_6 column figures by 0.8, and the F_{11} column figures by 0.6, adding the two corresponding products in each row to produce the new F_6 loading. This calculation is approximate only, except when the axes after the nth rotation are at right angles.

The steps may be summarized and illustrated as follows: First normalize (1, tan θ) where θ is the angle through which the RV is moved. If these values are called a and b, then the new projections, F_1', are $a \times f_1 + b \times f_2$, where $f_1 \times f_2$ are the corresponding projections on RV's F_1 and F_2, and where the sign of b becomes negative if F_1 was shifted *away* from F_2.

For example, in Fig. 1, Diagram 20, page 200, Chapter 12, where the tangent of the angle of rotation of F_1'' toward F_2'' was 0.9, we normalize (1, 0.9) and obtain (1/1.34, 0.9/1.34) or (0.74, 0.67). Now using the first two columns of the V_2 matrix, we obtain the following figures for F_1''' which compare closely with the first column of the V_3 matrix, page 202, Chapter 12.

TABLE 37.

Test	$.74F_1'' + .67F_2''$ =	Approx. F_1'''	True F_1'''
1	$(.74)\cdot(.36)+(.67)\cdot(-.48)$	$-.06$	$-.05$
2	$(.74)\cdot(.59)+(.67)\cdot(.05)$	.47	.47
3	$(.74)\cdot(.06)+(.67)\cdot(-.62)$	$-.37$	$-.37$
4	$(.74)\cdot(.92)+(.67)\cdot(.03)$	.70	.70
5	$(.74)\cdot(.60)+(.67)\cdot(.64)$	.87	.87
6	$(.74)\cdot(-.08)+(.67)\cdot(-.05)$	$-.09$	$-.09$
7	$(.74)\cdot(.08)+(.67)\cdot(.08)$	.11	.11
8	$(.74)\cdot(.07)+(.67)\cdot(.11)$	.13	.13

This method offers a saving on the usual method, working through the λ matrix, because one has to multiply out only two columns instead of all, to get one column in the V_n matrix. But neither this nor the graphical method soon to be described saves one having to calculate the λ and C matrices at each step. For without the C matrix one might allow two vectors to converge to very large intercorrelations or even collinearity without noticing it, and without the λ matrix being carried forward one cannot prevent the error becoming cumulative through reverting back at the end to a clean correlation from the unrotated matrix.

The arithmetical and graphical devices of this kind can therefore safely be used for only a few rotations at a time, after which the relatively approximate run must be checked by interpolating the usual type of calculation through λ. We shall therefore call these *methods of saving by intermittent runs.* The saving is not quite as great as it might be because the λ matrix, though not used, must be kept up to date, being recalculated at each shift. For there is no short general formula for transforming the λ_n matrix to the λ_{n+x} matrix; the x successive calculations, being all different, *must* be made in sequence. Since the other F's involved in shifting F_a alter from rotation to rotation, it is best to make the calculations while one is sure of dealing with the right F, i.e., the F_x of the correct generation, on which the shift was made. However, if one wishes, it *is* possible to store up the λ matrices and make a single calculation of the new RV position after such an intermittent run of, say, two or three shifts, by calculating the successive products of the individual shifts. This involves careful indexing to keep systematic records of the manner in which

those factors toward which shifts are to be made have themselves changed their positions. Thus, omitting for the moment the normalization calculations and considering instances where the factors shifted upon have not themselves shifted, the new angles after a run of three shifts would be:

$$[((F'_1+\tan\theta_2 F'_2)+\tan\theta_3 F'_3)+\tan\theta_4 F'_4]$$

where F'_1 shifts of θ_2, θ_3, and θ_4 were made, respectively, on factors F_2, F_3, and F_4, and if F'_1 represents the column of direction cosines of F'_1 on the unrotated factors.

But each set in parentheses above needs to be normalized in turn before being admitted to the larger bracket. Moreover F_3 may no longer be F'_3 by the time a shift is made on it. It may be F^2_3 and itself equal to $(F'_3+\tan\theta_x F_x)$.

Taking account of the normalization for a run of only three shifts the above expression becomes:

$$\left[\frac{\dfrac{F'_1+(\tan\theta_2)F'_2}{\sqrt{F^2_1+(\tan^2\theta_2)F'^2_2}}+(\tan\theta_3)F'_3}{\sqrt{\dfrac{(F'_1+(\tan\theta_2)F'_2)^2}{(F'_1)^2+(\tan^2\theta_2)(F'_2)^2}+(\tan^2\theta_3)(F'_3)^2}}+(\tan\theta_4)F'_4\right]$$

$$=\left[\frac{\dfrac{F'_1+(\tan\theta_2)F_2}{\sqrt{F^2_1+(\tan^2\theta_2)F^2_2}}+(\tan\theta_3)F'_3}{\sqrt{1+(\tan^2\theta_3)(F'_3)^2+\dfrac{2F'_1F'_2(\tan\theta_2)}{(F'_1)^2+(\tan^2\theta_2)(F'_2)^2}}}+(\tan\theta_4)F'_4\right]$$

which is already sufficiently cumbersome to persuade one to calculate each λ matrix afresh at each shift on all factors even though one does not use it at each shift for multiplying out V_0!

GRAPHICAL METHODS WITH MECHANICAL AIDS

The *graphical* method of intermediate runs, shortly to be described, can safely run from five to ten successive rotations without checking back to V_0—in the hands of a skilled person—and it then offers considerable saving. The apparatus required is a drawing board, a long T square with transparent stock, and a large set square (about 10-inch) also transparent. Each drawing instrument should have a special

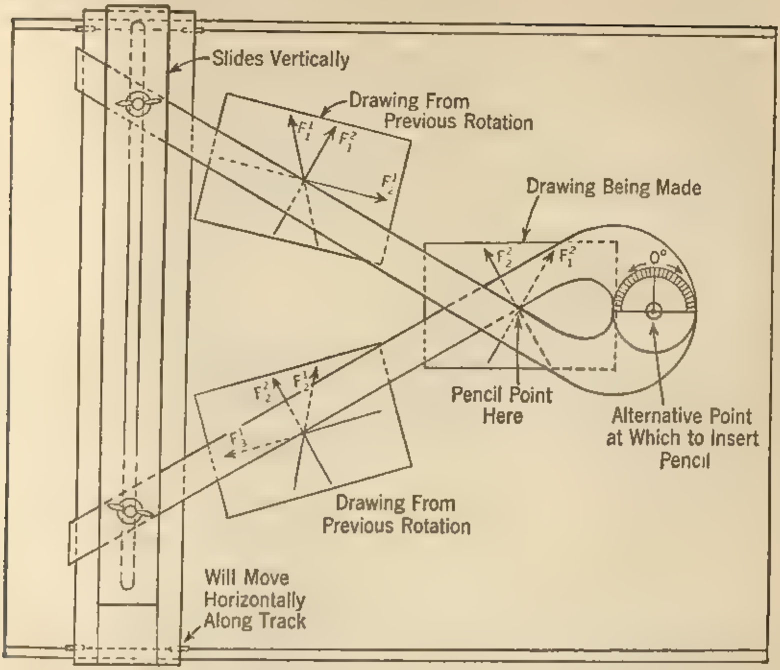


Fig. 2. Rotometer for Carrying Out Oblique Rotations Without Calculation.

DIAGRAM 26

guiding line engraved about 1/4 inch in from its working edge on the underside, so that one does not experience such distortion or parallax error as would result from working with the actual edges. At the right angle of the set square, where these lines intersect, a conical hole big enough to put a pencil point through is drilled through the set square (see Diagram 26). These special lines and hole can be made in a few minutes by anyone handy with tools.

The procedure, which can be demonstrated more quickly than described, is as follows: One draws in the new *RV*'s on the graphs, according to the indications of hyperplanes yielded by inspection, in the usual way. Let us suppose we have made such shifts for F_2 on F_1 and for F_5 on F_9 in Diagram 26, so that the variables (points) now have new projections on F'_2 and F'_5 though we have not calculated them. We now proceed to draw the next round of graphs without any intermediate calculation (except to calculate the new λ matrix

which is carried along but not used each time). This drawing is done for the $F'_2 \times F'_5$ graph, for example, as follows.

The old F_1F_2 graph is put at the left of the drawing board with the drawn-in F'_2 axis arranged vertically (by set square). At the same time the F_5F_9 graph is placed at the upper middle of the board with the F'_5 reference vector arranged horizontally (by T square). A new sheet of paper, which need not be graph paper, is lightly tacked to the board with Scotch tape directly on a level with the F_1F_2 paper and directly vertically below F_5F_9. With set square and T square one first marks the origin, O, on this $F'_2F'_5$ graph-to-be, dropping perpendiculars from the origins of the F_1F_2 and F_5F_9 graphs, with set square and T square, at the same time drawing the new $F'_2F'_5$ axes vertically and horizontally through this origin, as shown in Fig. 1, Diagram 26.

The variable points being numbered in the drawing, one begins with finding number 1 on F_1F_2 and F_5F_9, sliding T square and set square till it lies under the engraved lines on these, as shown for point p in Fig. 1, Diagram 26. At this juncture, pushing a pencil point through the drilled hole in the right-angle corner of the set square, one marks the position of the new point p on $F'_2F'_5$. (The reader may note that though the F_5F_9 projections will be directly under their true positions, those carried by the T square from F_1F_2 are displaced by the distances of the engraved lines from the edges of the instruments, but since this is consistent for all points, including origin and axis, no error, of course, results.)

This graphical intermittent runs method has on its debit side (1) that the points must be numbered when drawn (though this is often done in the ordinary method), (2) that combined shifts cannot be made, and (3) that one has to search for a point with a particular number in both graphs. In Zimmerman's more extended description of a similar method (**149**) which the reader may care to compare, the graphical method is said also to have the disadvantage of being restricted to orthogonal rotations. But as the method is developed here this is no longer true. For, in the first place, *slight* obliquenesses can be handled with the method of Fig. 1, Diagram 26 and where they cease to be slight they can be handled by the method shown in Fig. 2.

The apparatus shown in Fig. 2, and contingently called a Rotometer, can be readily made in any laboratory workshop. It consists essen-

tially of a couple of transparent plastic rulers coming together at a hinge on the right and having the angle between them fixed by butterfly nuts in the vertical slide at the left. The triangle thus formed has to retain its angles and its orientation on the board while being movable (by motion of translation) to any position.

One first places the drawing being made upon the board and draws upon it the angle between the given reference vectors indicated by the calculated C matrix. The angle between the rulers is then adjusted (by the fly nuts) until the edges lie along the hyperplanes of these RV's. The two previous drawings from which one is working are then tacked to the board with Scotch tape so that the hyperplanes of F_1^2 and F_2^2; i.e., the chosen rotations of F_1^1 and F_2^1, also lie along the (corresponding) edges of the rulers. The angle can be checked by the protractor on the hinge. The triangle is then shifted until the same numbered point, on the two drawings, lies on these edges. The position of the new point on the new drawing is then give by the intersection of the ruler edges and is marked by putting the pencil at the spot shown by "Pencil Point Here" on Fig. 2. The insertion of the remaining points requires only one movement of the triangle for each.

As the line of intersection of two thick rulers is sometimes such as to create slight parallax error, especially with an uneven pencil, an improvement of the apparatus consists of having a conical hole drilled through the joint as shown at "Alternative Point . . ." on the right of Fig. 2. The drawing being made is then placed farther to the right and its origin is fixed by placing a pencil point through the hole when the ruler edges lie on the origins of the other two drawings. An alternative construction of the rotometer apparatus is to run the vertical slide in a vertical slot in a draughtsman's planograph, attached to (an extension of) the top left corner of the board. This gives easier movement but not quite such rigid orientation.

Parenthetically it is possible to make a *make-shift* apparatus for oblique axis drawing by a slight addition to and modification of the common drawing materials used in Fig. 1. One takes an extra set square or small transparent oblong of about the same thickness of material as the T-square. One now arranges the graph at the top of the board more to the left if the angle is going to be drawn acute, and more to the right if it is going to be obtuse. The first reference vector is set at the proper angle to the vertical (the vector is set vertical as before). The only other innovation from Fig. 1 is that

the set square must now be made to run with one of its shorter edges perpendicular to the first *RV*. This is most readily done by having the second small set square or oblong of similar material, mentioned above, run in as a wedge between the large one and the T-square and temporarily attached to the latter by Scotch tape to give whatever angle is desired. However, the more rigid Rotometer is much preferable. Incidentally, it can also be used for orthogonal drawings by opening the rulers to the extent of 90°. For angles above 90° it

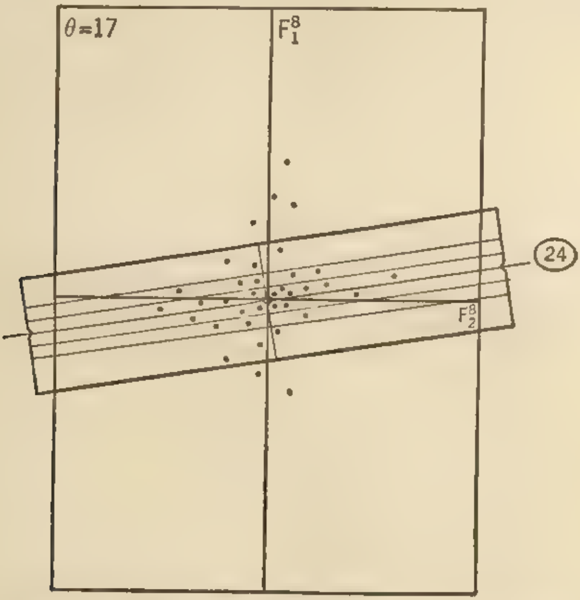


DIAGRAM 27. Rotascope for Judging and Recording Possible Hyperplanes.

is possible to shift the drawing around, though a Rotometer permitting the rulers to go into obtuse angles is simpler.

There remain one or two comments to be made regarding the gaining of impressions from graphs which apply to practically any methods of rotation. The use of a simple instrument, which has been called a rotascope, is quite an advance on the unaided eye in the matter of picking out hyperplanes. This, as shown in Diagram 27, is simply a transparent ruler with a hole bored in the middle and having an inner pair of parallel lines engraved in blue, and an outer pair in

red. The inner pair stand at a distance apart corresponding to ± 0.05 or ± 0.10 on the graph paper (the latter is probably better) and the outer ones at ± 0.15. The center hole is momentarily pinned on the origin of the graph by a pencil and the rotascope is moved around until a position is found where a maximum number of points, according to a count, falls between the parallel lines with which one is working at the time.

At that point pencil marks are made in channels centered at the ends of the rotascope and, if the rotascope can be so constructed, also in a groove through the origin, to mark the line and enable its tangent to be read, while the number of points is written at the upper end. When three or four drawings are finally selected and gathered together which show possible shifts on that particular *RV*, it will then be easy to choose, by a glance at these numbers, which is the best shift to take.

In the early stages of rotation, a wider hyperplane—say, ± 0.15—should be accepted. For the points that are eventually to settle in the true hyperplane are more widely dispersed in this proto-hyperplane. Indeed, at the very beginning it is often better to judge the hyperplane with the eye alone, apprehending the general sense of the swarm of points, for an actual count may then be a misleading penny-wise-and-pound-foolish argument. On the other hand, the eye is likely to be seduced by a relatively small number of points which happen in one drawing to form a neat straight line, when in fact an ill-shaped blotch of points may bring more into the hyperplane and into the neighborhood of the hyperplane as possible material to be brought in on some later rotation.

When a hyperplane appears at about 45° between the two reference vectors, it could equally easily be assigned to either. The choice must then be made according to which assignment would be more likely to improve the angles. In general we want to keep the inter-*RV* angles near orthogonality, and if one of the two *RV*'s is already sitting well in relation to most others, a large shift might ruin these relations. The choice must also be made according to which *RV* already has good alternative hyperplanes in other drawings. If F_x already has a hyperplane with 20 variables in a drawing waiting elsewhere, and F_y's alternatives rise to only a 15-variable hyperplane, then the present 18 variables hyperplane is best assigned to F_y.

This reminds us that there is no unique property right in the numbering of a factor (or rather, an RV). There is nothing but an historical continuity of calculation about the label of a reference vector; it is not mathematically the same for two rotations in succession, so there is nothing amiss about what might have been F_x becoming F_y. Indeed in the course of a long rotation factors may change places with one another, as far as meaning is concerned, without the rotator noticing it. The important thing is to keep the RV label, its λ_n, C_n, and V_n values all correctly associated. However, toward the end of a rotation process, and if the rotator knows the meaning of variables, it may give him additional guidance to recognize from past work that certain factors (as defined by their highly loaded variables) are likely to have a particular magnitude of variance and a particular degree of clarity or area in their hyperplanes.

Questions and Exercises

1. Why is it a mechanical aid to computation to set out the λ matrix in a transposed form? Describe the process of adding and normalizing the direction cosines from the old λ for the new λ matrix. What trigonometrical ratio of the angle of shift is used?
2. Discuss the conditions for good combined shifts and those under which such shifts are risky.
3. Discuss the pros and cons for making all possible drawings among factors and indicate some grounds on which choice of promising pairs, when not all drawings are to be made, can be based.
4. Using the factor matrix of question 2, Chapter 12, and the drawings of question 3, carry out a rotation to as good a simple structure as possible. Use numbered points so that combined shifts may be judged according to whether or not they will exert together more than the desired amount of change on the points to be brought into the various hyperplanes. In particular, would the following be good first shifts? Why?

On:	F_1	F_2	F_3	F_4	F_5
	$+\ .25\ F_2$	$-.10\ F_1$	$-.20\ F_1$	$-.120\ F_1$	$-\ .25\ F_1$
	$\pm\ .10\ F_3$	$\pm.30\ F_3$	$+.25\ F_2$	$+\ .40\ F_2$	$-\ .50\ F_2$
	$+1.00\ F_4$	$-.10\ F_4$	$-.25\ F_4$	$-\ .25\ F_3$	$+1.30\ F_3$
	$+\ .67\ F_5$	$-.60\ F_5$	$+.33\ F_5$	$+\ .20\ F_5$	$+\ .25\ F_4$

5. According to the method of rotating axes individually, would a shift on one factor ever improve the hyperplane of another factor? Why? What shifts with various axes, not mentioned above, might be better than those suggested, considering them as individual shifts? as combined with one or more other shifts?

6. Describe the method of intermittent runs using the direct factor matrix transformation. When does this introduce errors and why do these errors not vitiate the final simple structure?
7. Describe in detail the graphical method of making successive rotations without calculation of projections. Why is it necessary to carry along the λ matrix calculations, and why is this called a method of intermittent runs? What are its advantages and disadvantages?
8. Describe the use of the rotascope. What considerations other than number of variables influence the choice of a hyperplane position? How would one decide which RV to shift when the new hyperplane lies equally close to both of the existing tentative hyperplanes?

CHAPTER **16**

Other Specialized Techniques for Rotation

Having surveyed possible rotation methods and concentrated upon that which has the widest utility, we purpose now to give a little attention to some of the less widely used methods. For the problem of rotation to simple structure or to other criteria is still not solved by any universally best computing method; and the circumstances of a particular research, as well as the degree of skill in assistance and the availability of particular computing machines, are varying conditions which are best met by an intelligent adaptation of methods. However, the reader who is not yet at grips with problems in which he needs to consider finer points of rotation procedure and whose aim is to get a general technical view of factor analysis with adequate understanding only of basic rotation methods, is advised to skip this chapter and proceed to Part III, Chapter 17.

THE SINGLE PLANE METHOD

Most of the devices we can profitably consider are really aids to the main sectional view or single *RV* shift method discussed in the last chapter. However, there is one—and this we shall consider first—which is a true alternative rather than an ancillary. This is Thurstone's *single plane method* which, as indicated earlier, is not to be confused by reason of its label with the single graph shifts of the sectional view method. It is an alternative to be considered principally when a relatively mechanical procedure is required in the absence of experienced help, since it requires computational skill but no particular feeling for rotation problems.

So far the method of single plane rotation—thus named because one hyperplane (and its reference vector) is fixed before any others

are shifted at all—has merely been mentioned in the general survey of methods in Chapter 13. Compared with the sectional view method it is comparatively mechanical and blind, suffering from those dangers which can arise when a single plane is directed to its position independently of and without regard to the positions adopted by the other reference vectors. However, in practice, especially with the comparatively clear-cut structures found in abilities, it has proved quite satisfactory.

The procedure may be illustrated by the following example. From the factor matrix, select a test to be used as the trial reference vector and copy its factor loadings in the first column of a new table which we may call the redirection table, as shown in the upper part of Table 38. These are the direction *numbers* (they are not yet normalized) of the chosen vector, and in the example we have chosen test 3. We shall call these loadings a_1, a_2, a_3. In the second column of the table are recorded λ_1, λ_2, λ_3, the direction *cosines* of the test vector obtained by normalizing the a's, and these may be checked by the formula governing direction cosines of a line, i.e., $\lambda_1^2+\lambda_2^2+\lambda_3^2=1$. (It may be noted that so far this method simply follows the same procedure as that described in Chapter 12 for rotation by a leap to a trial vector position rather than by successive graphical shifts from an unrotated position. But in its later steps this is different from any other method yet described.)

We next find the projections of each of the test vectors upon our trial vector and record these projections in column 1 of a separate table, the *RV* matrix (trials record) at the bottom of Table 38. To compute these projections we multiply, as in matrix multiplication, the column λ_a (trial 1, redirection matrix) by successive rows of the original factor matrix adding the three products obtained for each row. In our example of eight variables, eight totals result to complete column 1. We now use these projections to plot against those of the original three unrotated factors, yielding three graphs, as shown in Diagram 28, upper row. All eight points may be plotted or alternatively only those which lie far enough from the origin so that they will be appreciably affected by a rotation.

We next draw a line through the origin on each drawing, as shown, such that it passes through, or near, as many of the points as possible that look as if they might conceivably eventually fall in a hyperplane. When there is a choice, the line forming the smallest angle with the

TABLE 38.

Redirection matrix

Test 3

		a	λ_a	s	$\lambda_a - s$	$s\lambda_a$	$1 - s\lambda_a$	$\epsilon = \frac{\lambda_a - s}{1 - \lambda_a s}$	λ_a'	
Trial I	F_1	.11	.18	.08	.10	.01	.99	.10	.10	$\Sigma a^2 = .3869$; $\sqrt{\Sigma a^2} = .62$
	F_2	−.38	−.61	.10	−.71	−.06	1.06	−.67	−.68	$\Sigma l^2 = .9638$; $\sqrt{\Sigma l^2} = .98$
	F_3	.48	.77	.12	.65	.09	.91	.71	.72	$\cos \alpha_1 = .9872$; $\sin \alpha_1 = .16$
Trial II	F_1		.10	.00	.10	.00	1.00	.10	.10	$\Sigma l^2 = 1.0254$; $\sqrt{\Sigma l^2} = 1.01$
	F_2		−.68	−.05	−.63	.03	.97	−.65	−.69	$\cos \alpha_2 = .9924$; $\sin \alpha_2 = .12$
	F_3		.72	−.10	.82	−.07	1.07	.77	.72	
Trial III	F_1		.10							
	F_2		−.64							
	F_3		.76							

RV (factor) matrix

	Original			Trials record		
	F_1	F_2	F_3	1	2	3
1	.33	−.48	.21	.52	.51	.50
2	.39	−.29	−.34	.00	−.01	−.03
3	.11	−.38	.48	.62	.62	.62
4	.91	−.20	−.34	.02	−.02	−.04

RV (factor) matrix

	Original			Trials record		
	F_1	F_2	F_3	1	2	3
5	.55	.24	−.65	−.55	−.58	−.59
6	.04	.07	.16	.09	.07	.01
7	.68	.47	.34	.10	−.01	.03
8	.60	.44	.31	.08	−.02	.01

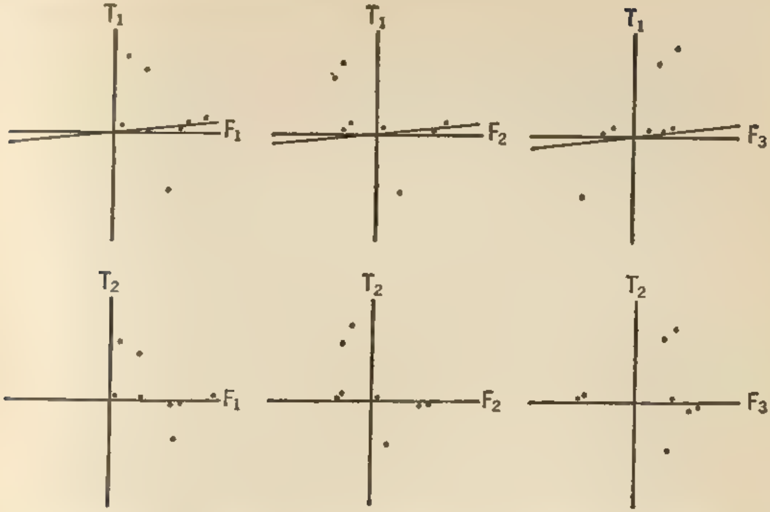


DIAGRAM 28. Graphs for the Single Plane Method of Rotation.

line at right angles to the trial vector should be used. Once selected, the slopes of these lines (tangents of the angles formed between each line and the corresponding factor axis) are measured and recorded in the third column of the redirection table (Table 38) headed s. With the value of λ_a and s now obtained we next carry out the steps indicated by the 4th, 5th, and 6th columns of the redirection table. For example, in the first F_1 row, $\lambda_a = 0.18$, $s = 0.08$. Then

$$\begin{aligned} \lambda_a - s &= 0.18 - 0.08 = 0.10 \\ s\lambda_a &= (0.18)(0.08) = 0.0144, \text{ or } 0.01 \\ 1 - s\lambda_a &= 1.00 - 0.01 = 0.99 \\ \epsilon &= \frac{\lambda_a - s}{1 - \lambda_a s} = 0.10/0.99 = 0.10 \end{aligned} \tag{27}$$

To fill in column 7, headed 1, we divide the entry in column 4, $(\lambda_a - s)$, by that in column 6, $(1 - s\lambda_a)$. Column 8, λ_a', is the result of normalizing the entries in column 7.

It is also found helpful in determining the progress of the rotation and the number of trials to be performed to record the magnitude of each shift made. To do this one takes the sine of the angle between the position of the trial vector at the beginning of any rotation and

its subsequent rotated position. The cosine of this angle is given by $\Sigma\lambda_a\lambda'_a$, and its sine is therefore $\sqrt{1-(\Sigma\lambda_a\lambda'_a)^2}$ since, for any angle θ, $\cos\theta=\sqrt{1-\sin^2\theta}$. When a rotation gives an angle between two successive positions of the vector, whose sine is in absolute value less than 0.10, that rotation may be considered the final one for this vector.

It may be seen then that this method has the effect of moving one trial vector and its hyperplane (here test 3) about while the other planes are fixed, in order to minimize as far as possible the distances of the original points from the hyperplane of the trial vector.

After completion of one rotation, another would be made with a chosen test vector orthogonal to the first, lying as nearly as possible in the plane just determined. That is, we would choose as our next reference test one whose entry in column 3 of the trials record, i. e., the last trial position, in Table 38 is small; in our example, number 2, 4, 6, 7, or 8. Similarly, the third test vector chosen would be one whose projection on both the first and second trial vectors, as finally settled, is small. This process may be continued, unless one is absolutely sure that the number of rotated factors is to be the same as the full number of unrotated factors, until no hyperplane can be determined for the reference vector chosen, at which point the rotation may be considered finished.

The objection to any method in which a set of variables has to be chosen arbitrarily, at a comparatively early stage of the game, to define a position to which the hyperplane shall be forced (as nearly as possible) by mathematical calculation is that the choice may be wrong! Instead of finding the true simple structure the process will merely tighten up and make more plausible the existing wrong guess. Also, as Horst (**75**) and Tucker (**130**) point out, most largely analytical methods tend to rotate toward the mean principal axis of the whole configuration instead of paying adequate attention to the specific subgroup concerned. This does not mean that the method is to be regarded as unsatisfactory, but that these dangers are to be recognized, and where a discrepancy arises between this method and the freer sectional view method the latter is to be regarded as offering the more likely solution.

Indeed having regard to general circumstances of computing assistance (skill, permanence, amount) in most psychological labora-

tories or sociology departments the writer is inclined to judge that the best rotation practice, at least in research of a basic nature, is the sectional view method as described earlier, aided by certain modifications now to be described. The single-plane method is preferable where the computing *has* to be made very mechanical, regardless of missing some finer features of the structure; the extended vector methods have an appeal where there are only four factors or so, while the semianalytical methods (Chapter 14) are quicker with very competent statistical help. But the sectional view method on the one hand does not require high statistical ability, and on the other is very sensitive to the real structure in the experimental material. It does, however, require circumstances in which the practitioner's experience can ripen, for any person long exposed to such rotation practice acquires certain artistic skills which both shorten the process and produce a finer product.

COMBINED ANALYTICAL AND GRAPHIC SOLUTIONS

The chief modification now to be recommended, apart from such devices as have already been suggested, is the borrowing of some processes from the semianalytical methods at strategic moments. It will be recalled that their essential innovation is the selection of a bunch of test points for the hyperplane which are then made to move into the hyperplane—or to scatter about it as closely as possible—by a single calculated shift. Now it can be objected to the claims of these latter methods to superior efficiency that if the bunch of tests that needs to be moved into the hyperplane is already as clear as has just been indicated, it is practically as simple to move them in by a graphical rotation. If the points lie radially from the origin as in Fig. 1 (Diagram 29), this is quite true; but sometimes a well-defined group lies nonradially as in Fig. 2. Here the process worked out by the semianalytical methods can be applied with great saving, for it would require several well-chosen shifts to bring these into the hyperplane by sectional rotation methods. The drawing of a graph, in supplementation of the analytical procedure is obviously desirable in such circumstances for the purpose of revealing which situation in fact exists.

Parenthetically it should be pointed out to the reader that in the shift in Fig. 2 there is no question of moving the origin of the reference axes, since such a move is not legitimate. Rather it is assumed

that in this case the points in the group are all above (or below) the plane of the paper, so that by raising the eye level (i.e., rotating F_1 toward us) the line of points can be seen against the line of F_2 (F_1's hyperplane section).

The adjustment here described involves, of course, a shift of the reference vector in question simultaneously upon all unrotated factors. The calculation of this movement can be based upon a value for the whole group of points or, if the group forms a well-defined line, upon values for a marker (or, better, a couple) at each end of the line. The second method, which is shorter, will be described here, the reader being referred to Thurstone (**126,** pages 377, 396) for the other.

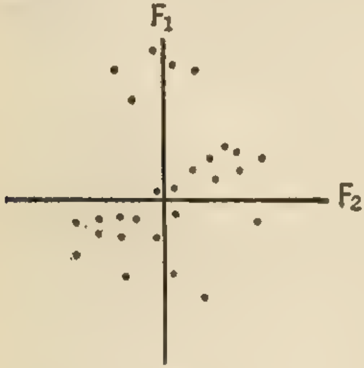


Fig. 1.

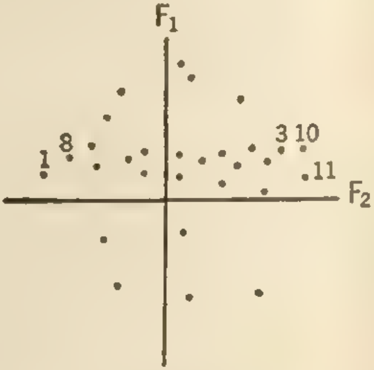


Fig. 2

DIAGRAM 29. Occasion for an Apparent Shift of the Origin.

Let us assume that we have taken a number of appropriate trial reference vectors using test vectors to guide us, as described above in the single-plane problem presented in Table 36. Let us next suppose that we have seen in the drawings a line of points that look as if they should go into the hyperplane (but cannot be rotated directly in) as in Fig. 2, Diagram 29. The process of shifting a line of points down into the hyperplane now requires that we first write the direction cosines of our reference vector (F_1) in the diagram with respect to the original unrotated axes. Let us use for illustration figures made up at random for a four-factor problem.

$$F_x = 0.33F_1 - 0.19F_2 - 0.57F_3 + 0.73F_4 \tag{28}$$

This is a column out of the λ matrix.

Next we pick out one or two (or more) points marking the ends of the line seen in the drawing, such as points 1 and 8, and points 3, 10, and 11 in Fig. 2, Diagram 29.

From the V_0 matrix we now calculate the *mean* of the projections of 1 and 8, and similarly for 3, 10, and 11, getting two sets of direction numbers as follows (the numbers are made up for this example):

$$C_{1,8}=1.11F_1+0.56F_2+0.87F_3-1.43F_4 \tag{29}$$
$$C_{3,10,11}=-0.40F_1-1.29F_2+0.75F_3+0.81F_4 \tag{30}$$

These are normalized, as explained earlier, and become direction cosines

$$U_{1,8}=0.53F_1+0.27F_2+0.42F_3-0.69F_4 \tag{31}$$
$$U_{3,10,11}=-0.23F_1-0.74F_2+0.43F_3+0.46F_4 \tag{32}$$

We now take the inner products (see page 207) of each end of the row with the direction cosines of the existing reference vector F'_x, thus:

$$F'_xU_{1,8}=(0.33)(0.53)+(-0.19)(0.27)+(-0.57)(0.42)+(0.73)(-0.69)=-0.62 \tag{33}$$

$$F'_xU_{3,10,11}=(0.33)(-0.23)+(-0.19)(-0.74)+(-0.57)(0.43)+(0.73)(0.46)=0.16 \tag{34}$$

and with one another, thus:

$$U_{1,8}U_{3,10,11}=(0.53)(-0.23)+(0.27)(-0.74)+(0.42)(0.43)+(-0.69)(0.46)=-0.46 \tag{35}$$

Thereby we obtain the cosines of the angles between the two vectors used and between them and the reference vector. These are needed to get certain constants in the following final solution formula.

Thurstone (**126,** pp. 377, 409) has shown that the required new reference vector, F''_x, is defined by the following formula:

$$F''_x=F'_x+p_1U_{1,8}+p_2U_{3,10,11}$$

where p_1 and p_2 are constants. This means that each of the four direction numbers of F'' are obtained by taking the corresponding direction cosines of F'_x, $U_{1,8}$ and $U_{3,10,11}$ and after multiplying each U component by an appropriate constant to be explained in the next paragrapn, adding the three resulting numbers to obtain one direction number.

To find p_1 and p_2, we first must compute the following values:

$$
\begin{aligned}
(U_{1,8}\times U_{3,10,11})^2&=(-0.46)^2=0.21\\
1-(U_{1,8}\times U_{3,10,11})^2&=1-0.21=0.79\\
(F'_x\times U_{1,8})(U_{1,8}\times U_{3,10,11})^2&=(-0.62)(-0.46)=0.29\\
(F'_x\times U_{3,10,11})(U_{1,8}\times U_{3,10,11})^2&=(-0.62)(0.16)=-0.07
\end{aligned}
\tag{36}
$$

Now we can calculate p_1 and p_2 as follows:

$$
\begin{aligned}
p_1&=\frac{(F'_x\,U_{3,10,11})\times(U_{1,8}U_{3,10,11})-(F'_x\,U_{1,8})}{1-(U_{1,8}U_{3,10,11})^2}=\frac{-0.07+0.62}{0.79}=\frac{0.55}{0.79}=0.70\\
p_2&=\frac{(F'_x\,U_{1,8})\times(U_{1,8}\times U_{3,10,11})-(F'_x\,U_{3,10,11})}{1-(U_{1,8}U_{3,10,11})^2}=\frac{0.29-0.16}{0.79}=\frac{0.13}{0.79}=0.16
\end{aligned}
\tag{37}
$$

With these last two values we are now able to compute the direction numbers of F''_x in relation to F_1, F_2, F_3, and F_4 as follows:

$$
\begin{aligned}
F''_{xF_1}&=0.33+(0.70)(0.53)+(0.16)(-0.23)=0.66\\
F''_{xF_2}&=-0.19+(0.70)(0.27)+(0.16)(-0.74)=-0.12\\
F''_{xF_3}&=-0.57+(0.70)(0.42)+(0.16)(0.43)=-0.21\\
F''_{xF_4}&=0.73+(0.70)(-0.69)+(0.16)(0.46)=0.37
\end{aligned}
\tag{38}
$$

Thus,

$$F''x=0.66F_1-0.12F_2-0.21F_3+0.37F_4 \tag{39}$$

which is our new reference vector.

Normalizing, we obtain the cosines for a unit vector:

$$=0.86F_1-0.16F_2-0.27F_3+0.42F_4 \tag{40}$$

which constitutes the required column in the final λ matrix.

OTHER SHORT CUTS IN GRAPHICAL ROTATION

A much simpler way of bringing a line of points crystallizing parallel to the hyperplane down into the hyperplane can sometimes be used, namely, on occasions when all the variables happen to be positively loaded on the first or centroid unrotated factor. A shift on the centroid can then be carried out simultaneously raising or reducing the loadings of all these variables on the given reference vector. This is achieved by increasing or decreasing (according to whether all the points need to be made more positive or more negative) the direction cosine (in the λ matrix) of the given RV to the

first unrotated factor. The adjustment can be made either by drawing the graph of the RV to F_1 and reading off, in the usual way, the tangent that will bring the variables into the RV hyperplane, or by a shrewd guess as to the increment required on the existing cosine (followed by renormalizing the λ column). Such a move on the centroid will not shift all the points equally, but it will shift them more uniformly than a move on any other factor would, providing the first centroid loadings are, as usual, large and positive. (Incidentally, the preservation of the possibility of making this shift is one reason for reflecting the original variables, as suggested on page 151, to keep all the first factor loadings positive.)

In general, indeed, when a reference vector has been drawn with every other in the current V matrix without finding any shift that will improve its unsatisfactory hyperplane, one must remember that it can still be drawn with every one of the *unrotated* factors. The unrotated matrix must always be in mind as a possible source of reinforcements when *in extremis.* The best unrotated factor for this purpose can generally be seen from inspection of its loading pattern in relation to the embryonic hyperplane of the RV in question. One can also fall back on earlier V matrices as a source of new RV directions in which to seek an effective move.

A shift as advocated in the last paragraph, based not so much on a drawing as on knowledge of the loading pattern that exists in one particular factor and the change of loading that is required in another, can be used with respect to other factors than the first centroid or indeed than any of the unrotated factors. As rotation proceeds one tends to become familiar with the projection patterns on the tentative RV's reached at any given stage of the rotation. For example, one may notice that through several shifts variables 1, 2, 7, and 40 have remained very high on RV number 1. At such a point of familiarity one sometimes becomes aware that the set of variables which are near but not in the hyperplane of another factor are these same 1, 2, 7, and 40 which happen to be high on RV one. A slight movement toward or away from the latter, i.e., the addition of a slight positive or negative fraction of its direction numbers to those of the factor, may then suffice to complete its hyperplane without any drawing having been made. Again, if a particular variable seems to move with a group but not quite to align itself with the hyperplane,

one can read the matrix row to find a factor on which it is high and take a very small shift on that factor as an ingredient in the next combined move.

FAMILIAR PHENOMENA IN THE ROTATION PROCESS

Finally, in rotation by graphs, we must take account of one or two frequently recurring events in the history of the diagrams which the experimenter needs to acquire ways of handling. First, it sometimes happens that an elliptical distribution appears between two *RV*'s, as in Diagram 30 below. A moderate ellipse of this kind may appear through drawing on orthogonal axes what is really a circular distribution on the oblique axes, in this case with an acute angle between F_1 and F_2 (Fig. 1, Diagram 30, true drawing in lower left). If, therefore, such a shape appears when the C matrix indicates r's of, say, 0.5 to 0.8 between the *RV*'s, the matter may be ignored (except perhaps that a redrawing with oblique coödinates is indicated for accuracy of rotation). But a lenslike shape occurring with approximately orthogonal coördinates (Fig. 2, Diagram 30) is more significant and indicates one of two things, either the presence of what have been called coöperative factors or the loss of a dimension.

Regarding the likelihood of loss of a dimension, it has already been stated that in general in the extraction process one should err on the side of extracting too many factors rather than too few, since the rotation process is capable of eliminating the excess. Let us now see how this occurs. It is essentially a result of looking for simple structure, for there is no reason whatever why mere rotation of the axis system, without some additional criterion, should lead to a reduction of the number of axes. The reduction occurs because one cannot find enough hyperplanes—at least, enough hyperplanes with any substantial factor variance left upon them. Thus one knows that one has taken out more factors than the actual structure of the data warrants and that some factors are mere error variance. The last of the unrotated factors may well be dropped on discovering this condition in the rotated factors.

The elimination of surplus factors in rotation occurs in two ways. First, we may find two *RV*'s approaching one and the same hyperplane. They fold up either by acquiring similar direction cosines and similar loadings or by showing an extremely narrow ellipse of points when plotted together, whether orthogonally or in true oblique ro-

tation. In the first case one can best proceed by averaging their direction cosines, calling the resulting *RV* the single inheritor of both. In the second an elliptical shape may appear, as stated above, *without* any obliqueness of axes. One can best proceed then by putting one hyperplane through the long axis of the ellipse and the other through the short. More frequently the latter will better pass through some clear hyperplane that emerges *roughly* across the shorter axis (as shown at F_1' in Diagram 30, Fig. 2). In this case F_2 has become almost a residual, losing most of its variance, and it will usually be found that on plotting it against some further factor, especially one of only moderate variance, it will give up the rest of its variance in

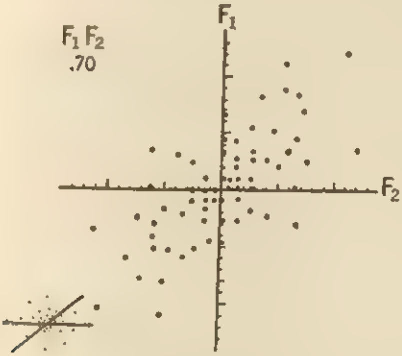


Fig. 1.

Fig. 2

DIAGRAM 30. Two Ways of Handling an Elliptical Distribution.

a similar elliptical disappearance. The factors of smallest variance often give these ellipses when plotted together, and thus reduce to negligible variance. The only drawback to taking out factors beyond the probable number and eliminating them in this way is that when too many are in excess and when the elimination is not carried out with thoroughness, i.e., when the residual carries away some appreciable variance on one or two variables, the patterns of the remaining factors are apt to suffer a little in accuracy through fragments of true variance being carried off in this debris.

COÖPERATIVE FACTORS

Some apparent elliptical plots, however, are, as stated above, coöperative factors and, as far as we yet know, do not indicate a need to eliminate. This phenomenon—so far found most clearly in the

DIAGRAM 31. An Actual Example, from Psychological Data, of Coöperative Factors. (From *Description and Measurement of Personality* by R. B. Cattell. Copyright 1946 by World Book Company. Reproduced by permission.)

realm of general personality factors—is characterized by plots yielding two satisfactory hyperplanes (and for approximately orthogonal factors) but with a concentration of the remaining points in two opposite quadrants, so that at a glance the distribution is not unlike an ellipse. A couple of real factors illustrating coöperation are shown in Diagram 31. Such a constellation means that there is some tendency for the variables that are highly positively affected by F_1 to be preferentially highly positively affected by F_2, and similarly for negative influences, though the two factors are nevertheless distinct and uncorrelated. If we seek for examples of such relations elsewhere in nature, we find in general, but not always, that some third factor can be discovered which is a function of the first two. As examples we can take cold and damp which affect a variety of human infirmities similarly (namely, adversely), yet which are distinct in nature and in

their influences on other variables; or latitude or altitude, which have similar influences on flora and snow precipitation and several other things, yet which could be separated in a factor analysis by the fact that some things, e.g., length of day, would lie in the hyperplane of one, being unaffected by altitude, and not in that of the other. Physiology yields a very clear instance of coöperative factors in the sympathetic and parasympathetic nervous systems which are responsible for individual variations of response in a large number of common variables, e.g., blood pressure, skin changes, endocrine secretions, but which are essentially distinct and which could be rotated to distinct positions by taking care to include enough variables affected by one and not by the other (**30**).

The phenomenon of coöperative factors has been little explored and it is possibly more widespread than we suppose. Whenever, in any large sample of variables, two factors show community of area regarding the variables they affect and a uniform sympathetic or opposed direction of influence upon them despite being distinct influences, this phenomenon of elliptical distribution through coöperation is likely to make its appearance. It is possible that in some research findings where incomplete factor extraction before rotation and rather rough rotation procedures have led us to suppose that only one factor exists (along the long axis of the ellipse) two coöperative factors remain to be distinguished in later research. Alternatively, it may happen that the choice between the interpretation in terms (a) of two coöperative factors or (b) of a single factor, the *RV* of which lies along the long axis of the ellipse, is really a choice between different kinds of explanation, as described in connection with the efficacy of factors (page 113). Thus, in the above instance of a supposed factorization of variables of flora, meteorology, and physical features for a population of positions at various altitudes and latitudes one possible interpretation of the variance is in terms of two coöperative factors—altitude and latitude. But another is in terms of a single reference vector along the ellipse—temperature.

One may notice further, however, that the addition of extra variables to the matrix might make possible the emergence of temperature as an *additional* factor to altitude and latitude, taking away much of the variance that was in these. Again, it is possible that temperature could be made to emerge as a second-order factor from first-order factors for altitude and latitude.

One source of coöperative factor structure appearing in certain instances is undoubtedly, as Saunders has pointed out, the appearance of two factors where one is literally some mathematical function of the other, especially some higher power of the other, so that they stand, for instance, as x to x^2. They then influence variables in a comparable fashion because they are variants of the same thing. For example, in a population of variables formed by variables and characteristics of sailboats we might find a factor for height of mast and another for sail area—the latter, for example, loading speed, angle of heeling, distance of visibility, thickness of running gear. Since sail area would be a rough second degree function of height of mast, these factors would be coöperative.

These considerations, germane as they are to further developments in the art of factor analysis, cannot be followed up further here. It suffices if the reader has perceived that bringing the rotation process to a proper conclusion involves something more than a mechanical pursuit of simple structure. There is required in this art an appreciation of the meaning of some of the constellations, e.g., coöperative factors, that may frequently be encountered and a sense of the meaning of cues to various reference vector movements, factor eliminations, etc. in relation to the ultimate meaning of the factor resolution that is being approached.

Questions and Exercises

1. How does the single-plane method of rotation differ from the sectional view method.
2. Using test 3 in the factor matrix of Chapter 12, carry out the single-plane method of rotation as described in this chapter, through one complete rotation with drawings. Are more rotations indicated by the size of sin ϕ? What is the reason for forming the columns λ-s, $s\lambda$, 1-$s\lambda$, and l?
3. When is the single-plane rotation best used to obtain simple structure?
4. In what circumstances are the semianalytic methods referred to in this chapter most helpful? Describe the essential steps in Thurstone's method of moving a single line of points into the hyperplane.
5. Under what circumstances is a shift on the first centroid likely to be helpful?
6. Discuss the desirability and the risks of error in extracting more factors than are likely to be required. What manifestations during the rotation usually accompany the disappearance of a superfluous dimension?
7. Describe the possibilities of making a shift of one factor upon another without making a drawing and without any analytical computations.
8. Describe the phenomenon of cooperative factors, discuss its possible causes and the indications which it may offer for the rotation process.

Part III

GENERAL PRINCIPLES AND PROBLEMS

CHAPTER 17

The Effects of Errors

As indicated in the preface, the purpose of Part III of this book is to proceed to matters which belong neither in the elementary general survey of the meaning of factor analysis in Part I nor in the guide to proficiency in the practical, technical processes of Part II. Part III is concerned either with wider theoretical vistas which also have some bearing on refinements of practical processes, or with special refinements of practical aids. Thus, although it appeals primarily to the reader wanting a more complete theoretical integration of the subject, it also has some necessary hints to the person who wants to develop the best possible devices for practical work.

So far we have passed lightly over the vexatious subject of errors and the way in which they may affect our ideal processes and conclusions in factor analysis. The point of most immediate practical relevance is their effect upon our judgment of how many factors to extract from a correlation matrix—an issue postponed until the present on account of the complication of its theoretical solution.

CLASSIFICATION OF ERRORS

Broadly viewed, errors may be divided into those which are essential, i.e., unavoidably part of scientific research by statistical methods, and those which are nonessential, i.e., appearing in particular studies but avoidable by some special pains. The former include two major groups (a) sampling errors, i.e., the tendency of any particular sample of persons[1] to have a different mean and sigma in

[1] In the widest sense, sampling error includes errors of sampling of variables, as well as of persons, as is obvious in Q-technique. But the concept of an ideal universe of variables is still so little discussed that for practical purposes we shall confine at least our quantitative formulations to errors of sampling of persons from the ideal complete population.

any measurement from that of the total or ideal population and (b) errors of measurement, in which we can include errors arising mainly from the experimenter, e.g., faulty observation, interpretation, scaling or recording, as well as ambiguous instructions inviting irrelevant responses, and errors on the part of the subject that are additional to function fluctuation, such as responding with disregard of instructions.

Errors of experimental measurement are recognized in the specification equation by adding a term e for each individual on each performance, thus:

$$P_{ij} = s_{1j}F_{1i} + s_{2j}F_{2i} + \dots \qquad s_j F_{ji} + e_{ji} \tag{41}$$

In their common separation from the common factors one often thinks of e_j and F_j (the specific factor) as a single block of irrelevance, and it is true that in many formulas error and specificity (uniqueness) can be put together as if they were one. For example, as far as prediction from common factors is concerned, the specificity and error are alike inaccuracy. Although they are distinct enough in our conceptions, we can separate them only by special experimental and statistical designs. For example, the reliability coefficient shows the agreement of the test with itself due to both common factors and specific factors, and falls short of unity only as a result of true error. The difference between the square of the reliability coefficient and the square of the communality thus gives the variance due to the specific factor.

Chance errors, if constant, i.e., added to *every* individual's score, and uncorrelated (as between tests) would not affect the correlation coefficients, but they are not constant; and though systematically they have no correlation, they do have chance correlation. The general effect of nonconstant errors, i.e., errors not affecting everyone similarly, is, as every student knows, to reduce the correlations by attenuation. It has been shown by Saunders (**107**) generalizing the finding of Roff (**104**) that attenuation errors will not systematically affect the existence and nature of simple structure or the number and general form of the factors, but the loadings of any variable in any factor will be systematically lowered to an extent given by the simple formula

$$\text{True loading} = \frac{\text{Calculated loading}}{\sqrt{\text{Reliability coefficient of variable in question}}} \tag{42}$$

The pattern of loadings of factors on a *variable* will thus not be changed by experimental error except to be lowered as a whole, but the pattern of loadings of variables by which a *factor* is recognized[2] will be much altered if tests have widely differing reliabilities or change their reliabilities, e.g., by changing length, as between two researches. This is important to keep in mind when comparing loading patterns as between two researches, and it is probable that much dispute on the identity and invariance of factors would be dispelled if factor matrices were published with corrections for attenuations. Incidentally, this can be done more economically by correcting the loadings after extraction than by correcting the r's for attenuation before extraction (see **104**), except when the factors are more than half as numerous as the r's.

Saunders also shows that the standard error of a loading (which has been corrected for attenuation) equals

$$\frac{(1-r_j^2)a_{jk}}{2r_j\sqrt{N}} \tag{43}$$

where a_{jk} is the loading of variable j on factor k, and r_j is the reliability coefficient of j. N, as usual, is the number of persons in the population. Thus the correction for error of measurement is dependent only on the reliability and the number of cases. (These of course will be the same when testing the significance of a difference between the loadings of the same variable on any two factors in the same analysis.)

SAMPLING ERROR AND THE NUMBER OF FACTORS

Turning now to sampling error, we find that its effect is in general, as Thomson (**120**) says, to "blur the outlines" of factors and make it easier apparently to fit the findings to *any* hypothesis—a fact to be kept in mind in discussions on factor invariance below. Experimental measurement error also blurs outlines, as it affects individual variables differently; but sampling error increases error in all correlations and also affects the variance of factors as wholes. Sampling errors can occur either in terms of univariate selection, i.e., the given sample is selected to differ from the population variance in some one measurement, as, say college students may be in having high intelli-

[2] The shape of the pattern by which it is precisely *recognized* will change but not the mere collection of variables by which it is marked out from the hyperplane and casually recognized.

gence, while being unselected in personality variables, or in terms of multivariate selection, where we suppose the scatter of the sample affects all variables. Since it is somewhat uncommon in natural samples to find them definitely screened with regard to one variable only, it is multivariate selection that deserves more attention.

Thurstone has shown (**126**) that univariate and multivariate selection (of a partial kind) do not alter simple structure, or the loading pattern of factors, but the size of the loadings is reduced[3] in proportion to the selection while the correlations (obliquity) among factors will also be modified. Thomson (**120**) reminds us that sampling errors, as opposed to experimental errors, tend to be systematically correlated, i.e., when we select a sample with regard to one performance, we generally find that it turns out to be selected also in regard to another. Consequently, he argues that sampling errors will not only enter into the isolated e term of the specification equation (41), p. 292, along with experimental error, but also will affect the number and nature of common factors. He argues, in opposition to Thurstone's original stand, that they may even *create* small common error factors out of what were part of the unique factor variance.

It seems that this possibility of errors being correlated must now be accepted. Just as we can get a correlation between two variables that is due to chance, so we can get a correlation among three, four, or more that is due to coincidences in the errors of measurement. Naturally the chances of getting such a correlation of appreciable magnitude become very small as the number of variables among which it is required to hold becomes of any size. But very small pervasive correlations may thus exist, and they will give rise to a slim common factor over the area of variables in question. These are too small to have any practical consequence except in the matter of deciding when to stop in the process of extracting successive factors from a matrix. To this practical problem we must now turn, fortified by the present discussion of the error problem.

As pointed out in our first encounter with this problem, in connection with communality estimation, mathematical discussions of the number of factors to take out of a matrix are likely to be preoccupied with fixing the rank of the matrix by some maximalizing

[3] The student will recognize that this follows the rule for ordinary r's which are reduced when the variability of measurements in either or both variables is reduced, i.e., if homogeneity of the sample is increased.

or minimizing rule. This approach is misleading in two respects. First, it tends to adopt an artificial goal. Since one can modify the rank of a matrix by making a particular choice of communalities for the diagonals, the mathematician tends to assume that one should deliberately modify it in accordance with some single principle such as obtaining a minimum rank, maximizing the specific variance, maximizing the number of common factors, etc. These are questionable; the important thing is to find the true rank as indicated by the best fit with the existing correlation coefficients, i.e., the off-diagonal *r*'s.

But here the second misleading value intrudes. The fact is that one is not interested, from a scientific point of view, in obtaining even the most natural rank of the matrix. The matrix is blurred by errors, some of them correlated to produce spurious factors, and what one eventually wants is the number of real factors, not the number of real plus error factors which is the rank of the matrix.

At first sight then, from the practical standpoint, it may seem that the unknown communalities are the chief obstacles to determining the true rank of the matrix, and certainly this has to be overcome before we begin wondering about the number of error factors. For, as indicated in Chapter 10's discussion of estimating communalities, we can, up to a point, make more common factors come out by estimating them to be large, and fewer by estimating them to be small. But this difficulty can be overcome, even though it has to be by laborious means. In the first place we can use the centroid analysis and practice iteration, i.e., repeat the analysis, inserting each time the better communalities obtained at the end of the last analysis. Or we can use Lawley's method of maximum likelihood referred to on page 396.

Then when the fitting of the most likely communalities (not the communalities to maximize this or that special feature) has been achieved, we face the question as to how many of the factors now extracted are real and how many are due to error. The unknown communalities, indeed, involve only what we have called above non-essential error. It is the real errors of measurement and sampling which leave us in doubt as to whether the last one or two factors of small variance, properly extracted according to mathematical standards, are real influences in the scientific phenomena or just error.

As indicated when comparing the various factor extraction methods (Chapter 11), some are more subject to the effects of errors than others. The group centroid extraction methods are more subject to the distortion of communalities by the accumulation of chance errors in the intercorrelations of the small group and this is especially true of the multiple group method, where all are extracted together and there is no chance to correct communality by watching the residuals from successive factors. Such error sources only reach dangerously distorting proportions in the multiple group method when unduly small groups are taken. An appreciable part of the observed intercorrelation among as few as four or five variables may actually be error, and the error variance thus carried into the factor space is not removed by subsequent iteration procedures. If the factorization seems to be producing far more factors than were expected, if the variance in these is high, and if the communalities of many variables come near to or exceed unity, it is probable that the computer is taking too small a fraction of the total variables into each cluster. Computers need to be fairly constantly reminded to take the more difficult course of searching far and wide for members of a cluster. The fact that a variable has entered into one cluster should not preclude its entry to one or two others; and, in general, from a third to a tenth of the variables (depending on the size and nature of the matrix) should be involved in any cluster.

TESTS OF COMPLETENESS OF FACTOR EXTRACTION

What we now want to know is when to stop extracting factors, i.e., at what point we have taken out enough factors to cover the true factor space. Or, if we take out real and error factors we want to know how many of the series to drop, in rotation or otherwise, as error factors. A great variety of approaches—theoretical and empirical—have been made to this; and since it is still not generally agreed which is best, and since different conditions presented to the experimenter will force him sometimes to take less than the best, we shall first list briefly any devices which have some degree of positive value. After this review of historical efforts we shall attempt to evaluate them.

1. The device adopted by the early workers in the field was that of comparing the standard deviation of the residual correlations, after taking out the supposed last factor, with the standard error

of the original correlations. But this does not work satisfactorily, for the residuals are similar to *partial* correlations (with factors held constant). The standard error of these residuals should really be divided by the uniqueness (the not-common-factor variance) to make it comparable (see 10 below) so that one would not expect the residuals already to be merely error at the point when the probable errors of the original r's are reached.

2. One of the useful early practices has been to plot the *distribution* of residual r's after extracting what is judged to be *almost* the right number of factors. This distribution should cease to be skewed and should approach normality when no more common factor variance exists, for there is no longer any systematic influence, namely, a factor, causing them to depart from chance distribution around zero. The method has not been widely tried out, and raises a second problem of deciding what *degree* of departure from normality of distribution must be considered incompatible with complete extraction.

3. Mosier (**97**) tried several methods, against particular factor problems having known factors and errors, and found some effectiveness in method 3 and the next two methods to be mentioned. His method in the first place seeks an indication that the standard deviation of residuals after the last factor has been extracted has become less than the standard error of the *mean* correlation in the original r matrix. Thus far it closely resembles 1 above, but it also asks that a plot of sigmas of successive residuals should flatten markedly after the last true factor is extracted. The first part of this is open to the same objections as 1 above, though it works roughly.

4. Another criterion is that when the product matrix is worked out for the factor in doubt, the maximum contribution to *any* r should be negligible, e.g., less than 0.10, and that the curve of mean contributions from these inner products should flatten markedly after the last real factor is extracted.

5. The maximum distance of the centroid from the origin, as measured by the maximum value (when all permitted reflections have been made) of

$$\Sigma_j^n \Sigma_k^n \rho_{jk}$$

where ρ is a residual r, and k and j are variables obtainable by reflection of the tests through the origin, would flatten (as plotted for successive factors) when the last real factor is extracted. Mosier

found this one of the best of the methods then tried, but the reader is referred to his article (**97**) for amplification.

6. Tucker (**125**) developed a criterion on a wide empirical basis which has been one of the most successful, in the writer's experience, among the very short methods, and which has recently been given some theoretical support. It utilizes the observations in the above three methods that various functions of the residual matrices or the product matrices, when plotted for the successive factors, tend to show a sharper deceleration of the drop after the extraction of the last true factor than at other points. The r's in the residuals are added without regard to sign, and include the communality residuals, giving a total which we will call Σl for the residual after the last factor. One more factor must be extracted beyond the supposed last in order that we may calculate also $\Sigma(l+1)$, the sum of the residuals after the $(l+1)$th factor. Then, according to Tucker's criterion, the expression $\sqrt{\Sigma(l+1)}/\sqrt{\Sigma l}$ equals or exceeds $(n-1)/n$ (where n is the number of variables) when l is really the last factor needing to be extracted.

In practice this value sometimes rises and falls instead of rising uniformly—indeed the sum of the residuals (but not the sum of squares of residuals) can itself start to rise again after the extraction of a certain number of factors. By this method it is unquestionably possible to get occasional absurd results, but its empiricism contains some intuitive truth which makes it the most reliable and practicable of the really quick tests.

7. Reyburn and Taylor (**102**) propose a criterion with a relatively simple theoretical basis but involving rather more work, namely, to divide the sigmas of each of the original r's into the corresponding residual r's, (the quotient should be unity on the initial naive hypothesis in 1 above). They then plot the distribution of these quotients and assume that if it departs significantly from normality, more factors are still to be extracted. This test belongs to the same genus as 3 above and 10 below.

8. Coombs (**39**) introduced the criterion of counting the number of negative signs left in the residual matrix after every possible variable reflection has been carried out (i.e., after every attempt has been made to reduce the negativeness of the matrix). Naturally this number will vary with the number of tests, and the following norms are consequently required. When the number of negative signs

remaining exceeds the figures given below, it is considered that all factors have been extracted.

a. Variables in matrix	10	15	20	30	40	50
b. Irremovable negative signs	31	79	149	358	660	1061
c. Standard error of b	5	7	10	15	20	25

It will be noticed that the standard error of the criterion number is so large, especially with small matrices, as to make application of this criterion necessarily a little rough.

It has been the experience alike of the writer and other users of this criterion that it leads to extraction of too few factors. Coombs anticipated this, saying (**39**) the "criterion does not indicate the point at which all the common factor variance has been removed, but rather it indicates the point at which the common factor variance remaining is overshadowed by the error variance remaining." This may be a useful point to recognize, but is not the best point at which to cease extraction. It is better to have all the factor outlines a little blurred by chance error than to have some factor variance totally missing, for if the latter occurs every rotated factor will be systematically missing substantial parts of itself which should have come from the later unrotated factors that have been omitted from the extraction.[4]

9. Swineford (see **71**) advocates correlating the original r's (strung out by unraveling the rows of the correlation matrix like wool from a sock) with the series of corresponding residual r's. When no trace of significant relationship remains, it is evident that the extraction must be complete. This does not appear to have been widely tried out.

10. McNemar (**94**) criticizes most criteria on the ground that they do not consider the numbers of subjects on which the r's are based, and he sets out to improve the theoretically faulty criterion based on the probable error of the original r's, as mentioned at the beginning of this list. This improvement consists in recognizing that the residual r's need to have their standard error calculated as if they were partial r's (the factors having been partialed out). The standard error σ_{res} that we really need to know is, he argues

$$= \frac{\sigma_r}{1 - M_{h^2}}$$

[4] It is a mistake, moreover, to assume that all the chance error is in the factors last extracted.

where σ_r is the observed standard deviation of the residuals after r factors have been extracted and $M_h{}^2$ is the mean of the communality for all factors (through the rth) on all the tests. ($1-M_h{}^2$, it will be observed, is a measure of the variance of unique factors.) When σ_{res} falls below $1/\sqrt{N}$ (N being the number of subjects) the extraction is considered complete.

This also tends in general experience to stop factorization too early, though not so much perhaps as in the Coombs criterion.

11. Saunders (**109**), in a comprehensive analysis involving consideration of the reliability of the measurements, the number of subjects, the type of correlation employed, and the variance of the residual factor, has arrived at a test which starts from the same firm theoretical basis as McNemar's, but which he claims is an advance thereon in that it attends in the formula not only to the number of subjects in the population, the reliabilities and the number of variables, but also to the number of factors extracted. It takes the following two alternative forms which are theoretically similar but which start from a different basis of calculation

$$\sum_{ij} {}_k\rho_{ij}^2 < \left(\frac{n-k}{n}\right)^2 (n-\Sigma_i h_i^2)^2 \frac{1}{N} \tag{44}$$

$$\sum_{ij} {}_k\rho_{ij}^2 < \left(\frac{n-k}{n}\right)^2 (n-\Sigma_i r_i)^2 \frac{1}{N} \tag{45}$$

where i is one variable and j another
k is the (order) number of the factor being extracted.
n is the number of variables
${}_k\rho$ is the residual correlation after the kth factor extraction
N is the size of the population
r_i is the reliability coefficient of the variable i.

The expression on the left is the obtained sum of the squares of all the residuals in the matrix after the kth factor has been removed. It is this which is being tested against the criterion on the right which has been computed as indicated in the footnote. When it falls below this criterion, the extraction is complete. In the first form the criterion on the right starts off with the sum of all the communalities of the factors extracted up to the point, whereas the other starts off with

the reliabilities of the variables when they happen to be known.[5] The second is slightly more convenient and should be used where the reliabilities of the variables are known. As Saunders shows (**109**) these criteria give reasonably good, but not always exact, answers when tried on synthetic examples where the number of factors involved is known for certain. Naturally an absolutely exact answer is not to be expected, for the essential errors which we have discussed are only to be estimated by a statistical likelihood, and in any given case may be higher or lower than the probable error.

CHOICE OF CRITERIA

Of the above criteria the theoretically and practically most effective is probably Saunders', while the shortest is Tucker's. If one wishes to determine the end point of factorization with the utmost possible accuracy—though such accuracy is unimportant and seldom attempted in the present exploratory use of factor analysis in social science—it is best to use a *combination* of criteria, since none is infallible. Thus one might take Saunders' criterion, with checks from those of Coombs, Swineford, and Tucker. If necessary one could, further, take the communalities from the number of factors thus decided upon and refactor with Lawley's maximum likelihood method.

However, in the majority of researches it must suffice to check by only one, or at most two, criteria. The choice has to be dictated by available time (criteria 1 and 5 can be most quickly applied; 4, 6, 8,

USE OF SAUNDERS' CRITERION

[5] For the nonmathematical reader the steps required for the test by this criterion may be briefly indicated as follows:

Square and add up the residuals after the kth factor. If the diagonals (communality residuals) are not counted,—and it is generally more accurate *not* to include them unless there is reason to believe in exceptionally good communality estimation—the total must be expanded by multiplying by $2n/(n-1)$ to bring it to that equivalent to a complete matrix. Call this A, set it aside and calculate the criterion as shown in the formula as follows:

a. Divide the difference between the number of variables and the number of factors so far extracted by the number of variables and square the result. Call this B.
b. Take the unrotated matrix, square all the loadings up to and including the kth factor loadings, add them (all $k \times n$ of them). Take this sum from n and square the result again. Divide the result by the number in the population. Call this C.
c. If A is now found to be less than $B \times C$, the factorization is deemed complete. If not, take out another factor and repeat this procedure.

and 10 require most labor) and by the extraction method used (the group methods, for example, do not apply easily to criteria 2, 3, 4, 5, and 7). After all, in most researches at present one is concerned to discover the *factor patterns* of important factors, and determination of the exact number of factors is important only in so far as inclusion or omission of the last real factor is likely to modify the form of some earlier factors in the rotation process. One is not actually concerned at losing some factor of very small variance (which can probably be picked up more strongly in some other matrix anyway). Nor is there any great point in knowing exactly the number of factors in a particular matrix *per se*—a transient, artificial collection of tests which may never again be put together.

SIGNIFICANCE OF FACTORS

Finally, although for practical purposes we must know when to stop factoring, i.e., when further extraction is bringing in no more real variance but only error, we should note that theoretically and in a wider sense the whole notion that a correlation matrix obtained from complex natural phenomena contains an *exactly* limited number of factors is incorrect. For the number of common influences, or basic personality source traits, which affects performance in any situation is almost certainly, from common sense considerations, quite large. By analogy, it is not a question of determining how many ships there are in a convoy but how many are visible from a given lighthouse and can be adequately illuminated. At the edge of the horizon it becomes an arbitrary matter as to whether one more ship shall or shall not be counted visible. Like optical devices in the convoy situation, the mathematical devices in the matrix will magnify or blur to include one or two more or less than the number substantially apparent; but strictly the common factors *actually at work to some very faint extent* (in any widely chosen, large matrix of variables) will exceed what we choose to call the rank of the matrix.

One thing is certain on both theoretical, scientific, and practical computing grounds—that it pays to extract too many rather than too few factors, and that the majority of criteria tend systematically to underestimate slightly the number of factors. On scientific grounds, as we have seen, the number of common influences among complex, personality and social indexes, and in the biological sciences generally, is large. Consequently the specific or unique factors are merely res-

ervoirs of unknown common factors not to be taken seriously as unique influences, and to be reduced whenever possible into common factors (by addition of new variables which share something with them) or error. On computing grounds the extraction of an extra factor or two can add very little to the error, for although error *may* be a larger fraction of the last than the first factor, the variance has become so small that it is unlikely to impart error of any magnitude. At the same time the true patterns of the factors are better reproduced than when one tries to rotate the factors in fewer dimensions than they are actually meant to occupy and without necessary elements of variance. The possible exception to this injunction to take out "too many" factors occurs in the use of the multigroup method of extraction, where there is a danger of building up one's groups from too few variables and of carrying in an undue amount of error variance—such as can enter into the r's of small clusters—into the common factor space. Proper use of the method will avoid this, and in such circumstances the plan of trying to take out a factor or two extra is still to be preferred.

This point could be illustrated by the failure of several important psychological studies to reach congruent scientific conclusions through premature arrest of extraction, but it is naturally most convincingly demonstrated by examples where factors have actually been put together in an artificial problem constructed behind the scenes and where the known number of factors can be compared with the factors extracted from the correlations by someone who did not know the underlying structure. One of the best instances of such an experiment is that of Mosier (**97**) who presented a known 4-factor problem with 20 variables and allowed computers to take out 3, 4, and 6 factors. The root mean square discrepancies of the obtained loadings from the real loadings (on the first three or four factors) were respectively 0.185, 0.064 and 0.053. Taking out more than the true number of factors thus gave more accurate results after rotation than when proceeding on a correct estimate of the rank of the matrix! In general practice, therefore, it is best to proceed to the point where the criterion indicates enough, and then to take out one or two more (depending on the patience of the investigator and the reasons for believing the criterion to be effective). Rotation will then eliminate or reduce to an obscure residual any factors that are in excess, by means indicated in the previous chapter.

This discussion of the significance of real factor variance left in a correlation matrix after extraction of so many factors may well terminate with the general question of a *test of significance* for any single (rotated or unrotated) factor. As Saunders shows (**109**) such a test takes the form:

$$\chi^2 = N\,\frac{n-1}{2n}\left(\Sigma_i \frac{a_i^2}{u_i^2}\right)^2 \tag{46}$$

where the symbols have the meaning stated earlier, a is a factor loading, and u is a measure of uniqueness,[6] i.e., $(\sqrt{1-h^2})$. Even though a test for the correct numbers of factors extracted has been applied beforehand it is valuable to apply this to any factor which has become very small and questionable in the course of rotation. (The degrees of freedom of χ^2 are $(n-k+1)$ where the k^{th} factor is involved.)

All the above tests of significance require the assumption that scores on the variables are normally distributed, although, as explained below (page 326), no such assumption is made for the computation of product moment r's or for the essential processes of the factor analysis itself.

RECOGNIZING THE SAME FACTOR IN DIFFERENT ANALYSES

It is appropriate in this general survey of the role of errors to take up as the next most important practical consequences of these theorems the technical criteria for identifying a factor in one research with that in another. This is cognate with the problem frequently referred to as that of establishing factor invariance. The question of identifying or cross-matching factors arises, of course, only when we are working with factors as real functional unities in nature which show themselves now in this context of research and now in that—for there is no question of matching two factors from the same matrix and experiment. When we have two distinct experiments, we should expect to recognize, in different batteries having all or a sufficient number of variables in common, the same patterns of loadings, which could therefore be ascribed to the same influences operating in a different sample of persons (or occasions, and partly of tests) and slightly different

[6] The reader may note that some statisticians symbolize the uniqueness by u and some by u^2, in various formulas repeated here. The present formula uses u^2, making u parallel to h, the basis of the communality.

conditions. Indeed, it has been taken as one criterion of the reality or efficacy of a factor, i.e., its existence as a constant influence, that it should remain invariant, i.e., having the same loading pattern, in the factor matrices of different experiments. Any such attempts at matching obviously require a proper appreciation of the role of error in distorting the true patterns.

Ample reasons have been given earlier (page 123) for believing that invariance will be attainable only when the results of the different studies are, in the first place, rotated to simple structure or to a unique, meaningful position based on some similar, definite, scientific criterion. One would expect invariance to fail when *unrotated* matrices are compared or to the extent that rotation is faulty. Nevertheless, claims have been made—quite incorrectly, the writer believes—to the effect that unrotated, bipolar factor solutions are also truly invariant. Incidentally, the fact that such claims to invariance of unrotated or falsely rotated factors can be made calls our attention to the important fact that when factor saturations of important tests are high, so that the test structure sticks out very boldly in space, quite poor rotational procedures—and sometimes even none at all—may fail to obliterate the essential patterns of high loadings that exist with simple structure. Indeed, the first one or two factors in order of variance are likely to be recognizably the same in a rotated matrix, a bipolar matrix and a principal components solution. But already in the second factor of unrotated, compared with rotated, solutions extraneous loadings can be seen creeping in, and by the third and fourth factors, any real resemblance in the pairs is generally lost. A more sustained parallelism exists between the series of unrotated centroid and principal components factors. But the search for invariance is meaningless unless the factorial analyses are aiming at the same general system of constellation (page 137) and unless rotation is used in the two experiments to be compared.

However, researchers are entitled to believe in the possibilities of obtaining invariance with *any* systems, whether they are likely to correspond to real functional unities or not, until experience disproves the fact. And whether this question of rival systems is at issue or whether we are only interested to identify the same factor in different experiments, we must have some *test* of loading similarity. Unfortunately, despite the great importance of such a test for the effective application of factor analysis in scientific research, we still have only

relatively untried devices to offer. These are the obvious devices which spring to mind, namely, (a) correlating the loading profile of the two factors to be compared; (b) comparing the profiles (loading patterns) by χ^2 or some derivative of it such as the coefficient of pattern similarity (**26**) which will indicate not only whether the profiles have the same *shape*, as does r, but also whether they are similar in *level*; and (c) dividing the variables into two categories according to the magnitude of their loadings, namely into those which mark a factor by being very highly loaded and those which have no significant loading, or alternatively, which significantly *lack* any loadings and fall in the hyperplane. In this last device—(c)—a second experiment can then be compared with the first to see how many of the markers and nonmarkers again fall in their proper categories and what the probability is that the obtained degree of sorting into these two categories could have been reached by chance alone.

Before evaluating these devices, one must bear in mind the observations made earlier in this chapter on the effects of essential errors upon factor loadings. Briefly, errors of measurement, by attenuating the original r's, will lower loadings; sampling errors will affect the relative variance of factors and their correlations one with another. If the errors of measurement (reliability coefficients) changed uniformly for all tests,—as would happen if one battery had tests just half the length of the corresponding tests in the other—all the factor loadings found in one research would be uniformly reduced in the other. Then the correlations of loading columns in one V_n matrix with those in the other V_n would reveal which factors are essentially the same in profile. But such uniform proportionality of reliabilities can rarely be guaranteed and it is in any case not safe to assume that a factor with loading pattern similar to, but lower than, another is the same factor. Examples are already known where two distinct factors have similar profiles differing only in level.[7]

Consequently, it is clearly desirable before applying criteria of identity of loading pattern to make corrections of loadings for attenuation (using the reliability coefficients). If the angles of factors are

[7] This appears in some matrices where the surgency pattern is paralleled by a similar factor, as yet unidentified, of much weaker variance. Of course, it is theoretically possible for two distinct factors to have the same pattern in both shape *and* level—providing none of the variables exceeds a loading of 0.72, i.e., $\sqrt{0.50}$, for the sum of the variance could not exceed unity.

also to be used as identifying characteristics, it is necessary to correct these for sampling influences (see **126**). When this is done, the coefficient of pattern similarity (**26**) clearly becomes a better device by which to test the similarity of the patterns than does ordinary r, for it tests agreement of *level* as well as shape. No expression has yet been worked out to show how far the coefficient may fall short of unity and still be accepted as satisfactory proof of identity in relation to the magnitude of chance errors existing; a test of whether the agreement exceeds chance could theoretically be derived from Saunders' expression for the standard error of a factor loading (page 293).

MATCHING BY COINCIDENT MARKERS

The third matching device described above, which does not apply measures of pattern similarity but simply puts variables into two categories—in or out—with respect to each factor, we may call *the method of coincident markers*. It is probably less sensitive than the *pattern coefficient* method, since it omits some of the available evidence, but its brevity recommends it for certain occasions. As worked out so far (**29**), it can be applied either to matching a single pair of factors or to finding the goodness of match simultaneously with respect to all factors in a pair of series based on two similar batteries. The problem may best be illustrated by a numerical example. In a set of 36 variables we may choose to mark each factor by the highest 6 in the loading for that factor. On comparing the markers with those of a second experiment with the same 36 variables, we are unlikely to find exactly the same 6 variables at the head of the factor considered to match the first, but we may find 4 or 5 of the 6 to be identical. How frequently might 4 or 5 out of 6 markers be the same by chance?

If we take one particular factor (say that of highest variance) in the first experiment and test its match with one particular predetermined factor in the second, the chance of such agreement is small. But in fact we do not generally have such predetermined factors to compare (at least until we come to the last unmatched factor on each side). Actually we take any factor among the dozen or so on one side that shows the best resemblance to any one of a dozen or so on the other and seek to test the goodness of that resemblance. Generally, therefore there is not much point in bothering with the expression for a single factor match where the factors to be compared have not been

otherwise predetermined.[8] For more frequently we are concerned to compare the series from two experiments *as a whole.* In that case, it will be noticed that we cannot speak of *exactly* 4 (or 5) markers matching, since on some factors it may be more and on others less, so that the most generally useful formulation would be in terms of *number of factors considered matched in the two studies*, a match being set at some defined level, say *3 or more* markers in common out of 6.

In general, the probability of *at least m* matches in two series of factors of n length equals

$$P_{N(M\ \ldots\ )} = \sum_{n=M}^{N} p^n(1-p)^{N-n}\ {}_NC_M \tag{47}$$

where p is the probability of success (obtaining a match between factors) and $(1-p)$ the probability of failure. ${}_NC_M$ is the symbol indicating the number of possible combinations of N things taken M at a time, and is also the coefficient of the Mth term in the binomial expansion of $(p+q)^N$ from which this formula is taken.

That is, if we let $q=1-p$, then $(p+q)^N=1$ no matter what power we choose to set equal to N. This is equivalent to saying that in N chances or tries some proportion of the tries will be successes and some will be failures, and if we consider every possible result and add the probabilities of all these results together, the result should be 1. (In other words, it is certain we will obtain a result if we do not care what it is.)

Now, if we stipulate that we must have *at least M* successes, then $(N-M)$ or less will be failures. We then add the probabilities of M, $M+1$, $M+2$, ... $N-1$, N successes, each of which is represented by a term of the binomial expansion $(p+q)^N$, and obtain:

$$\left[\sum_{n=M}^{N} p^n(\underbrace{1-p}_{=q})^{M-n}\ {}_{N_1}C_n\right] = {}_{N_1}C_M \cdot p^M q^{N-M} + {}_{N_1}C_{M+1} \cdot p^{M+1} q^{N-(M+1)} + \cdots$$
$$+ {}_{N_1}C_{N-1} \cdot p^{N-1} q + {}_{N_1}C_N \cdot p^N q^0 \tag{48}$$

[8] As it happens, it is necessary to work out this individual probability on the way to a general expression for matching the whole series in two experiments. Actually the probability for an exact 3 match out of 6 is comparatively simple, namely ${}_{33}C_3 \times {}_{30}C_6 = 0.0028$ if the two factors are already predetermined. If one is given and the other can be chosen from among 12 at random, it is twelve times this. In general, therefore, where each factor is marked by the h highest variables out of a total of t, and c of these occur among the h markers of any one of the n factors in a parallel experiment, the likelihood of such a match by chance is ${}_{t-c}C_c \times {}_tC_h \times n$.

The student who wishes to consider the probabilities of marker matches in more detail may do so in a tentative article elsewhere (**29**). Incidentally, it will be noted that all three of the methods for testing factor invariance and most of the concepts concerned therewith apply *mutatis mutandis* to testing instead the constancy of loadings pattern for a given variable, as expressed in the specification equation. In both cases we deal with the V_n factor matrix, but in the factors we read down the columns and in the tests we read the rows. However, changes in test reliability are more likely to upset the factor pattern and changes in sampling to affect the specification equation pattern.

Establishing the invariance of factors—their identity in different studies—is the more important problem, since, after all, there is rarely any need to identify tests. We *know* them by their explicit characters as our independent, controlled variables. Identification of factors need not depend on the matching of patterns or landmarks alone, but can also employ evidence indicated incidentally above, principally: (a) the nature of their correlations in the C matrix with other known factors; (b) the relative magnitude of their mean variance, by which we refer to the mean (or the root mean) of the squared loadings of the factor in all variables in the given study; and (c) proof by elimination. This is the circumstantial evidence that if all the other factors have been matched, in a study with the same set of variables, the remaining important factor on one side is likely to be the same as the remaining substantial factor on the other. Items (a) and (c) naturally depend on the field being fairly familiar, with most of the other factors already known to a high degree of confidence. To objectify and quantify the comparisons made in (a) and (b), the r_p coefficient of pattern similarity (**26**), as stated, is probably preferable to r.

Frequently, it will happen that the batteries in the two researches between which comparisons are being made do not contain exactly the same variables but possess only a large portion or a majority of variables in common. In that case an additional check on factor matching is obtainable by a new experiment in which the variables *not* common to the two matrices are put in a single battery to see if they keep company as expected from their associates in the separate factor matrices. This matter is taken up in the next chapter in connection with dovetailing matrices.

Incidentally, it is surprising, with these matching tests now available, that no research has been reported to evaluate the hotly disputed claims of various rotation methods to give greater "invariance" or stability for the factors found by them.

THE ROLE OF NONESSENTIAL ERRORS

Since the systematic listing of error sources at the opening of the chapter, nothing further has been said about what were called the *nonessential* errors, namely those arising from the guessing of communalities, the use of correlations with some measures missing, the incompleteness of factor extraction, and the presence of computing errors, especially rounding errors. On communality enough has been said to show that no complete solution exists. Apart from Lawley's method, which is too laborious for general use, or the process of iteration of the factor extraction which is seldom employed even in the least forbidding situation created by the multifactor extraction method, this error is generally uneliminated. It shows itself sometimes to the extent of some variables having communalities greater than unity and in failure of the inner products of the V_n matrix accurately to reproduce the correlations. A general idea of the magnitude of this nonessential error may be gained from a typical example. In the 20-variable, 4-factor, example of Mosier quoted above, errors in estimating communalities produced a root mean square error of 0.057 in loadings, while one of the essential errors—the chance error in the correlation coefficients due to sampling—produced a corresponding error of only 0.011. With larger matrices and the methods of communality estimation here described, however, the mean error of loading would be substantially less—perhaps of the order 0.02.

As to the second source of nonessential error just mentioned—that due to correlations based on slightly different samples through gaps in the data—perhaps the best thing to remember is that it *is* nonessential and avoidable by dropping individuals from the population until every one of those remaining has a score on every test. But we must recognize that it is the rule rather than the exception for the experimenter to find at the end of data gathering that some scores are either based on faulty responses or are missing—despite the most careful attention to attendance, scoring, and instructions. In a large project these gaps may be scattered so evenly in the score matrix, like moth holes in an otherwise excellent garment, that the

experimenter would be forced to throw away *most* of his results if he rejected every person with a score defective on some one test.

On the other hand, he may find that he can get as many as 80% of his total sample as a basis for each single r if he simply strikes out the defective cases in each paired series on which an r is based. He can argue then that since his total experimental sample is only a small fraction of the whole population of which it is a sample, he is not introducing much error by dropping 20% of such a sample, i.e., that an r on 80% is practically as reliable an estimate of the true r as one on 100% of his sample.

This is substantially true, but it is not the whole story. If Smith is missing on variable x and the correlation of x with y is considerable, the obtained correlation will be reduced by Smith's omission if he is an extreme individual, and raised if he is average. It can be shown by practical instances that correlating with missing cases in any substantial frequency tends to increase the number of common factors and to raise the communality of some variables substantially above unity. Only the barest minimum of missing entries can therefore be tolerated in a factor analysis.

The effects of rounding errors are at once too simple and too intricate a subject to discuss here—simple because the principles are well known to every schoolboy learning decimals, and intricate because the practice to be recommended depends too intimately on each method and case to be generalized. Most factor extractions do best to proceed with product and residual matrices kept correct to three decimal places, though in the last few factors one may shorten to two. Rotations can be carried out with single-figure matrix entries in the early stages, since there is no accumulation of error and they can be tightened to two-place accuracy as soon as the clarity of the structure requires more precise definition of the hyperplane. It would be hard to find any existing study in which the experimental and statistical conditions justify expressing the final rotated matrix with an accuracy beyond two decimal places.

Questions and Exercises

1. What are meant by the terms essential errors and nonessential errors and how is one of these entered in the specification equation?
2. What is meant by the standard error of a factor loading? Write the formula for it. If two analyses of substantially the same general data (but obtained from two different experiments) differ considerably in their

factor loadings, what types of errors in one or both of them might be suspected?

3. Describe what is meant by univariate and multivariate selection and indicate what is known about the effects of such sampling error upon the factorization.
4. Describe the effect of general sampling error in creating error factors. What enters into the calculation of the standard error of a factor loading and the test of significance of a factor?
5. Discuss the problem of finding the rank of a correlation matrix. Indicate some special guiding principles used in influencing this rank and show why finding the rank of the matrix does not suffice to tell the number of influences at work in a given scientific problem.
6. List and describe very briefly ten methods employed to ascertain when enough factors have been extracted from a correlation matrix. Are any of these superior to others? If so, in what ways?
7. What are some of the means of determining whether a factor obtained from one study is the same as a factor obtained in another similar study? Describe three criteria which can be employed upon the factor loading data alone.
8. Discuss the nature and the consequences of the chief nonessential errors in factor analysis.

CHAPTER 18

True Factor Resolution and Design of Experiment

A stage has now been reached at which we can penetrate to those more subtle issues about the nature and significance of factors which need to be faced in the process of factor resolution when alternatives and ambiguities present themselves. Such issues of factor meaning were faced in a general way in Chapter 8 at the conclusion of Part I, but it was not possible to handle them with insight and precision without the grasp of technical points achieved in the intervening chapters.

FACTORS AS SCIENTIFIC ENTITIES

So we approach again, at a new level, certain problems concerning the nature of factors, the recognition of spurious factors and artifacts, the relation of second and higher order factors, the dependence of factor structure upon some features of experimental design, and such phenomena as factor fission.

It will be understood that our discussion now takes place within the realm of factors obtained under conditions which give them more than merely mathematical reality. For though we began by discussing the infinitely variable, mutually transformable, possible sets of factors, obtainable from diverse systems of factor extraction and from resolutions directed to quite different ultimate factor constellations, we have now settled on the most convenient computational device and the most scientifically meaningful resolution. These are the centroid method or its derivatives, and rotation to simple structure permitting a constellation of overlapping group or general factors, supplemented when necessary by specifics.

In approaching general ,issues of experimental design we should

point out that this concentration on certain methods dictated by the needs of scientific investigation should not and does not lead us to overlook the virtues of other designs for special, restricted purposes—such as mathematical convenience or logical categorization of variables in a single matrix. Such objectives are sometimes proper phases in basic scientific research or remain as constant practices in special applied work. For example, for purposes of *classifying* a large number of variables in the most useful number of descriptive divisions (e.g., as in the tree of Porphyry **11**) regardless of *functional* unity and in relation only to what exists in the small universe of the *given population* of variables, the bifactor or bipolar factor constellations are, as Burt claims (**11**), especially apt. For another purpose—that of giving the most complete prediction with few factors or of reproducing the scores with greatest exactness—the principal components solution, beloved of the mathematician, is more efficient. It is rather difficult, however, to point to scientific purposes where these advantages in some special circumstance are not outweighed by the pervasive disadvantages of working with factors that are merely mathematical artifacts not possessing any other consequences.

But even when we follow a general system which aims at real, invariant factors corresponding to functional unities in nature, we still find, as our discussion in Chapter 8 showed, that factors may be said to differ in degree of efficacy. That is to say, some patterns may be found to repeat themselves in many circumstances of analysis, e.g., in both R- and P-techniques, as well as by behaving as experimental wholes or emerging as functions of some demonstrable physiological or sociological entity; while others are of lower efficacy because they are more restricted, conditional, or transient. We also saw that the same happenings may be ordered and conceptualized with equal correctness at times in two alternative sets of factor structures of equal efficacy, just as we may hold alternative perspectives visually in the same drawing. But all factors that are not merely peculiar to a single matrix, i.e., as mathematical factors only, have sufficient efficacy to be of scientific interest.

Let us look more closely at the first of these problems—that concerned with the varying degrees of efficacy of factors. We may approach it first by an attempt at philosophical, logical analysis of meaning, and second by an actual survey of empirically known factors attempting to relate their degrees of efficacy to their known natures.

LEVELS OF UNITARINESS

An adequate discussion of efficacy and the meaning of the unitariness which can be ascribed to factors would develop into a treatise on philosophy and logic which, though requisite to the fullest possible understanding of this scientific method, cannot be undertaken here. The essential discussion centers upon the meaning of unitariness, which has been thoroughly discussed by philosophers, but which does not appear to have reached a more definitive formulation than in the earlier discussion of Aristotle (**1**). Our thesis here is that the varying degrees of efficacy in factors, defined as the frequency of the diverse situations in which the factor loading retains its form, correspond to degrees of unity in the entity which the factor evidences.

It must be stressed, however, that the factor itself is not a unity; it is only the evidence of a unity. A factor, like any single variable, is a dimension or attribute of something. It is something measured in units in a single direction of continuous change. But it must be a dimension of something with some wide manifestation of its unity for a whole set of variables responds to changes in the measured dimension.

Without attempting to summarize exhaustively philosophical discussion on the nature of unity, we can yet state that degrees of unity are recognizable, and instance the following examples of such degrees.

1. Unity of conception in which the parts belong together only in the mind of the observer and not through any intrinsic natural properties, as when a woman thinks of a shopping trip, considering the particular stores she will visit as part of one trip. This comes near to being an accidental unity, the lowest form of unity, in which a collection of objects just happens momentarily to be together in space and time (Aristotle's first sense).

2. Unity of attributes. This comprises both Aristotle's similarity in kind and belonging to a common species. Here there is no common movement, as in 3, and the unity may even degenerate to that of 1, as when I pick out all the red objects in my field of vision, when the redness is so irrelevant to their nature as to give them no higher unity. Usually, however, common attributes will bring also some degree of common fate and thus give to the group a true, natural, operational unity. For example, if I throw iron and lead fragments into the sea, the former will soon be gone by rust, whereas the latter will remain.

A classical instance of a dispute as to whether a factor implies a unity of this order, or of 3 below, was that between Thomson and Spearman over the meaning of the general factor in abilities. Spearman asserted the unity to consist of a single energy, a mass action of the cortex in Lashley's sense, which came into action in many different situations. Thomson, on the other hand, considered the unity to consist of the existence of a limited collection of possible response connections from which varying numbers of items were selected according to the complexity of the task (**119**). The correlation of performances, as accounted for by the general factor, was therefore due only to the overlap of these various samples when chosen from a necessarily limited total collection. This unity of a reservoir of responses can be brought under the heading of unity of attributes, because these particular items have the attribute that they can be chosen, i.e., are suitable for, the application in question.

Incidentally the unity of "availability" in a single pool of possible responses, as involved in Thomson's concept of general ability, has come up again in a constellation of factors (see page 137) propounded by Guttman, too recently for inclusion in our systematic map of possible constellations, but in any case too specific to be appropriately included there. Guttman agrees with the position of this book that the real number of common factors operative in a situation is almost always very large, and he proposes a factorization which will make them as numerous as the tests. But there is in his method a peculiar relation among the numbers in the various tests, as follows (where F_c and F_s are respectively common and specific factors and T_1, T_2, etc., are test variables).

$$T_1 = F_{c1} + \ldots\ldots\ldots\ldots + F_{s1}$$
$$T_2 = F_{c1} + F_{c2} \ldots\ldots\ldots + F_{s2}$$
$$T_3 = F_{c1} + F_{c2} + F_{c3} \ldots\ldots + F_{s3}$$
$$T_n = F_{c1} + F_{c2} + F_{c3} + \ldots F_{cn} + F_{sn}$$

This differs from the other constellations (page 139) in involving *order* among the tests. He shows that for the simple progression shown a high ridge would arise along the diagonal of the correlation matrix; but there could also be more complicated cyclical orders which would give correlation matrices very similar to Spearman's hierarchy, and offer an explanation thereof by sampling, very similar to Thomson's. This theoretically ingenious analysis seems unlikely to correspond to

most natural structures. For example, the regular order of factor complexity above is unlikely, because in a region with n possible influences to sample from a test with $\frac{n}{2}$ common factors at work is far more likely to occur than one with n factors or one factor; i.e., there would be replications in the middle of the series above, breaking its simple arithmetical progress—the single representation of each combination. The proof of the pudding being the eating, it remains to be seen if correlation matrices are found in nature instancing this peculiar constellation. If so, for reasons which space prevents discussing further here, we should favor the explanation that such factors arise by overlapping pools of availability in a large number of "contributors to variance," each of approximately equal importance. That is, such factors are not likely to be different "entities" but differences within a large group in the "attribute" of availability.

3. Unity of systematic connectedness in space and time, as in the parts of a box or the syllables in a word. These appear together and disappear together. The unity of being created together, or historical unity, as well as the unity of growth belong here. This is essentially the unitariness indicated by correlation, at the level of the simple surface trait.

4. An intermediate between this invariable common fate or systematic connectedness unity and the highest organic unity is that in which in addition to common fate there is the lowest degree of interaction of parts, as when something done to one part affects all. Thus pulling one link in a chain affects all, but stopping one of a shower of raindrops does not affect the others. A shower is a unity of order 3, but a chain of order 4. The distinction between 3 and 5 is that between the philosopher's undivided unities and indivisible unities, and 4 lies between, in that division is *possible* but nevertheless produces serious changes.

5. An organic, integrated unity. Here, in addition to the characteristics of lower order unities there is interaction among the parts of such an order that the *whole* is powerfully modified and generally destroyed if one *part* is destroyed. A brick wall may have a section of bricks removed and remain a (shortened) brick wall; but if we remove the liver from a mammalian organism it is profoundly altered or destroyed.

EMPIRICAL ILLUSTRATION OF FACTOR UNITIES

To which of these levels does the unity of a factor usually correspond? Answering this can be assisted by the second possible approach to factor meaning, namely the empirical one. From Thurstone's box problem (**126**), where length, breadth, and height appeared as factors, and from William's factorization of the variables presented by instrument readings in a pilot's cockpit, which yielded altitude, compass direction, and speed, it is clear that factors can sometimes literally be dimensions. Other studies show that they can be metaphorical dimensions such as temperature or time. This seems to be a unity of order 2—of a common attribute.

Other instances of factors show structural or at least substantial entities, as in the sense of a chemical or hormone producing in various concentrations certain effects. We have already referred to an instance in which the factor could be the concentration of a common real element—as when in a variety of old wives' remedies for heart trouble William Withering (1785) noticed, by a process essentially of factor analysis without statistical aids, that digitalis (foxglove) was the invariable constituent and that the effect was strongest where the loading with digitalis was strongest (**36**).

Instances where the factor corresponds to an organic unity are admittedly hard to find, though general intelligence itself may prove to be a factor corresponding to the organic unity of the cortex. Possibly factorizations of sociological and historical data will provide instances where a factorial influence turns out to be a dimension associated with a person or an organically developing group, institution, or social movement. Our absence of sufficient examples to explore this question must be ascribed to two unfortunate circumstances. First, within psychology itself, in which most work has been done, too many investigators have been content to use factors without asking what they are. As Thurstone, (**126**), Stephenson (**116**) and Rimoldi (**103**) for example, have independently stressed, more attention needs to be given to inspecting mental processes in the performances highly loaded in a factor, to see what may logically be argued to be common to the performances.

A moderate amount of psychological analysis ancillary to factor analysis has been made in the field of abilities, both by expert psychometrists and by amateurs when their attention has been called in some

striking way to the nature of the underlying factor or influence accounting for correlatable individual differences. Thus one gifted amateur, Edgar Allan Poe, in the *Murders in the Rue Morgue* tells as "In chess the . . . chances of . . . oversights are multiplied . . . and it is the more concentrative rather than the more acute player who conquers. In draughts (checkers) on the contrary, where moves are unique . . . the probabilities of inadvertence are diminished, and . . . advantages are gained by superior acumen."

As no factor analysis of performance in indoor games seems yet to have been done, the question of whether distinct factors of alertness and intelligence, corresponding to this psychological analysis, can be demonstrated, remains to be seen. But the procedure of matching mental process with factor has been carried out with tolerable certainty in regard to the processes of spatial and numerical thinking in Thurstone's *S* and *N* factors and the dimension of familiarity with vocabulary and grammatical usage in *V* factor. Again, the hypothesis that Spearman's general ability factor shows itself in capacity to perceive higher-order relationships (analogies, classifications, etc.) in all kinds of material has been widely accepted. Rimoldi's recent analysis (**103**) of reasoning processes seems also to have brought out new alignments of first- and second-order factors with particular mental functions. In these instances the factor unity hypothesized is that of order 3—a systematic connectedness of the parts because they belong to a single mental process which either operates or does not operate, as a whole.

While the interpretation of personality factors is necessarily more speculative, due to the later development of factorization there, it also promises to provide illustrations of wider possible natural forms corresponding to factors. Some factors seem to express an amount of energy, a single power, while others seem to correspond to organic functional unities, in fact to the ego and superego structures long discussed as clinical entities by the psychoanalysts. (Factors *C* and *G* (**22**) have been claimed as the two factors matching these.) Factor *F*, or surgency-desurgency, with its happy-go-lucky, carefree placidity and conversion-hysteric symptoms at the surgent pole, has definite physiological associations, e.g., alkalinity of saliva, high skin resistance, low cholinesterase concentration (**23, 141**), which suggest that the factor may eventually turn out to be some single chemical pacemaker concerned with neural conduction. Again, other factors found in

dynamic responses, clearly correspond to the basic drives, so long speculated about in clinical and comparative psychology.

The second main reason why an insufficient empirical basis exists as yet for discussing the various natural correlates of factors is that research by this method has been largely confined to psychology, where the bases of observation tend in any case to be confined to subtle processes and abstract patterns. If more findings existed for review in sociology, economics, biology, meteorology, astronomy, and especially the physical sciences, this general theoretical question of factor meaning could be adequately treated. In a recent factorization of culture patterns (**27**) factors were found corresponding to wealth, population, size, degree of long-circuiting of behavior etc.; while in a factorization of historical data in the U.S.A. and Britain (**32**), factors were found corresponding to trends, e.g., to degree of industrialization. These indicate factors corresponding both to abstractions and to concrete influences and, on the whole, to higher orders of unity than in the psychological data.

SPURIOUS FACTORS

At present it seems that we can only assume that an increased order of unitariness will show itself by increased efficacy in the already defined sense of preserving an unmistakable if not invariant factor pattern through diverse factor-experimental designs, e.g., through R- and P-techniques and through factorization of increments, and in extraneous, independent experimental situations. By these criteria and others applied above our conclusion will be that most factors at present known seem to correspond to a unitary existence of order two, three, or four. Thurstone (**126**) has been content to say that "factors may be called . . . 'causes,' 'faculties,' 'parameters,' 'functional unities,' 'abilities' or 'independent measurements.'" But we have seen that they may vary considerably from abstract dimensions to functional, systematically connected unities such as drives or an emotional response pattern, and so to actual physical entities such as a physiological substance or a living organism. In the light of this broader conceptualization some of the automatically repeated warnings about the dangers of reifying factors are seen to be rather philosophically naive. One is quite as much entitled to reify a factor, i.e., to reify a factor obtained under the special conditions of simple structure, etc., as to use a sub-

stantive for the North Pole, a mechanical force, the French academy, the cost of living, or the family next door.

So far we have been speaking of degrees of efficacy or reality in the functional unities corresponding to factors which appear in rotated matrices and which show at least *some* degree of invariance in different situations, and we have distinguished these from mere mathematical factors. The researches of a number of careful, statistical workers have shown, however, that an intermediate or new species of factor also exists. It is in fact possible to run into factors which, though not merely mathematical factors, present only in one matrix and one arbitrary rotation, are yet so transient and tied to artificial circumstances as to be, for most purposes, merely confusing intrusions. Chief among these ghost factors are the factors produced by artificial conditions and spurious correlations arising from the form of the tests or the populations used. For example, if we have an item with four alternate responses in a multiple choice test and treat each (present or absent) as a variable in itself, we shall find four spurious factors each corresponding to an appreciable positive loading in one and a low negative loading in the remaining three—because the choice of one automatically reduces score on the totals corresponding to the other three.

Test and experimental design will generally do best to avoid situations where any relation is *bound* to appear between scores for purely mathematical or physical reasons in the test presentation itself. One must also beware of situations where various ratios are used as variables, some elements in the numerator or denominator of the ratios being common to several of them. For example, if from a timed test we take one variable which represents the fraction of attempted answers that the person gets wrong and another the time he takes to answer each item, a spurious correlation is likely to arise between them because "number attempted" is common to the denominator of both ratios. This is likely to issue in a spurious doublet factor. In general, the caveats which the student has learned for avoiding spurious correlation will also be useful in avoiding spurious factors.

DIFFICULTY FACTORS

But even when such gross artifacts are avoided, one may nevertheless run into a curiosity of factor analysis, the phenomenon of difficulty factors. It was shown initially by Ferguson (**51**) and by various

investigators that in correlating either single items or subtests, more than one factor will be obtained, when they range over a wide degree of difficulty, in situations where only one factor would be found if the tests have approximately equal difficulty or had a narrow total range. The phenomenon of difficulty factors arises in this particular form only when one test is very easy and another is very difficult, i.e., not when 5% pass one test and 5% pass another; but when, say, 95% pass one and 5% pass the other.

Wherry and Gaylord (**137**) showed that the appearance or non-appearance of difficulty factors is theoretically a consequence of the type of correlation coefficient used. They argue that though both the phi coefficient (or four-point coefficient)[1] and the tetrachoric correlation are derivatives of the product moment correlation, the former gives these spuriously enhanced correlations between items of very different difficulty whereas the latter does not. Recent research by Robert Smith and Demaree (**111**) indicates that this difference is only one of degree and suggests that the tetrachoric may produce slight difficulty variances akin to those in the factorizations of the phi coefficient and the product moment *r*.

While these facts are unquestionable, their interpretation remains open to further psychological discussion. In the first place, we must

[1] The ϕ coefficient is a simple derivative of the product moment correlation for use in situations where the facts give us only the percentages above and below a certain pass score. It takes the form:

$$\phi = \frac{\alpha\delta - \beta\gamma}{\sqrt{p_i p_j q_i q_j}}$$

where $p_i = \alpha + \beta$, and $p_j = \alpha + \gamma$ and are the percentages passing on tests i and j, respectively; and $q_i = \gamma + \delta$ and $q_j = \beta + \delta$, which are the corresponding percentages of failures. α is the percentage passing on both i and j;

β = percentage passing i but failing j;
γ = percentage failing i but passing j; and
δ = percentage failing both i and j.

	j	$-j$	
i	α	β	p_i
$-i$	γ	δ	q_i
	p_j	q_j	

For classes that are true dichotomies, i.e., not slices from continuous distributions, ϕ is better written

$$\frac{(bc)-(ad)}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

where a, b, c, and d are the actual frequencies in each of the four categories. But this value needs to be divided by a k value appropriate to the fraction in the largest passing class in order to be rendered comparable with r or tetrachoric r if the dichotomies turn out to be continuous instead of true dichotomies.

notice that these notions have emerged in pencil and paper tests mainly in education where it is usual to think in terms of a dichotomy of content and difficulty. Actually this dichotomy belongs to a rather subjective and naive stage of psychological thought in which it was supposed (as Guttman does in his scale analysis) that the homogeneity of content of a test can be decided by inspection. (In the case of a test of information this is plausible, but in personality tests it is an almost meaningless assumption, while in a collection of variables in, say, sociological or physiological realms, it is absurd to suppose that the experimenter knows immediately what is homogeneous or uniform in meaning. Since the object of factor analysis is to discover what is homogeneous, i.e., functionally unitary, the experimenter's prior attempt to put together items that are of homogeneous content can be regarded as nothing but an approximation directed by a hunch. If it succeeds, it saves some waste of items; if it fails, the factor loadings obtained still enable us to put together a fair number of items that should go together.

The very term content is unfortunate, for, as indicated in the earlier general discussion, the *meaning* of a test or variable does not lie in its subject matter or even in the form of the performance which it dictates, but rather in functional connections revealed by factorization and rooted *in the relation between the population and the test situation* (**22**) (or, in P-technique, the relation between an individual and the test). The content of a test can be given a useful meaning other than the loose popular meaning only as its factor content, i.e., in terms of its factorial homogeneity (or complexity), in terms of the specification equation. One cannot lock up a printed test in a closet and assume that in this alone one is storing a defined psychological trait. The trait meaning remains embedded in a set of relations definable by both the population and the environmental circumstances, which include the test. It would therefore be necessary to lock up the population too.

The dichotomy glibly drawn as to content and difficulty is therefore false in the sense used in these earlier writings. Actually there is nothing elusive about difficulty itself; the difficulty of a test can of course be operationally defined for a given population by the percentage failing it. But failing is the wrong term to apply to this performance, and the term difficulty is consequently misleading. The test may be one of, say, sociability vs. unsociability, where one direction of

performance is no more difficult than the other, and this is true of the majority of variables in social and biological science. The notion of failure in fact dissolves into nothing more than lying to one side of a score line arbitrarily fixed by the experimenter. Probably *eccentricity* would therefore be a better term than difficulty in indicating how far a given line in the test is from a 50/50 division of a typical population.

TEST CONTENT AND TEST ECCENTRICITY

As we know from our general study of univariate and multivariate selection of population, a large scatter in the population will sometimes increase the variance of the main factor among the variables and reduce the insignificance of certain other factors. Now increasing the scatter of a population while keeping tests fixed is equivalent to reducing the range of test eccentricity (difficulty) while keeping the population with a fixed scatter. Conversely a reduced range of population or an increased range of eccentricity—where tests on the face of things are largely of one factor—will increase the significance of the lesser factors while decreasing the variance of that large, usually general factor which has previously elbowed out the others. Such effects may be illustrated by the change in loading pattern in a number of formboard performance scores when measured first on children ranging widely in age and ability and then on a selected set of adults. The general ability factor diminishes and various group factors negligible in the first situation now become paramount. Again it has been shown by Guilford (**58**) that even in performances of the purest content, according to inspection, new factors appear as the eccentricity increases. Thus, auditory pitch discrimination breaks down into different but entirely psychologically meaningful factors at different levels of the performance (**58**).

On this basis some so-called difficulty factors (those in which the eccentricity is all in one direction) may be accounted for by the ordinary rules about magnification of variance of subordinate factors with change in variance on the majority of variables—as discussed in Chapter 19. But there remain eccentricity factors which are not genuine factors blown up from insignificant variance to significant variance in this way, but which, in accordance with Ferguson's first assertion (**51**) are artifacts appearing only with a certain type of correlation coefficient. Lawley (**84**), from a statistician's standpoint,

has followed Ferguson in indicating how these eccentricity factors can appear as mathematical entities occasioned by uneven difficulty in items or tests. That such factors depend for their very existence on defects in the designs of various indices of association has recently been empirically demonstrated and clarified by Smith and Demaree (**111**). They show that the trouble arises from the use of a coefficient (the phi coefficient and, in some circumstances, other derivatives of product moment r) which *cannot achieve unit covariance however good the correlation when the points of dichotomous division for the two variables differ considerably*, i.e., are extreme in eccentricity, in opposite directions. This is a fault of the coefficient itself, and Smith has shown that it can be at least substantially corrected by dividing each coefficient by *the highest coefficient obtainable* at that degree of eccentricity of the cuts on the tests. For example, when only 5% pass one test and 95% pass the other, if we suppose that 100% of those who pass the first also pass the second, we still obtain a coefficient of only 0.9474. Consequently, the coefficient for a *less* than perfect overlap between those passing test 1 and those passing test 2—say 80%—would need to be divided by this perfect value to show how it really stands. Thus,

$$[(0.04)(0.91)-(0.04)(0.01)]/0.0475=$$
$$0.758\text{, which divided by } 0.9474 \text{ becomes } 0.800.$$

It can be shown that when this is done, pure eccentricity factors disappear; and that the obtained factors when rotated all become psychologically meaningful (**111**).

Carroll's reëxamination (**150**) of Guilford's factorization of pitch discrimination, referred to elsewhere, showed as above that the correlations become systematically lower with increases in the eccentricity (difficulty) of items. But he demonstrated a fresh approach to correcting for this when he estimated the number of right and wrong responses from guessing in such forced choice binary responses, and showed that the differences from the "ideal" factorization largely disappeared when these were allowed for. Those who are compelled to work in factorization with the questionnaire, item-by-item kind of test, as in education (rather than with physiological or objective-type personality tests using a smooth continuum of response) would always do well to give attention to the effects of (1) degree of eccentricity (difficulty) and (2) the effect of the number of alterna-

tive responses per item. The first can be corrected as indicated above and the second by applying the usual correction for guessing (and an instruction in the test invariably to guess).

EFFECT OF FORM OF CORRELATION INDEX

It behooves us, therefore, in extreme populations or where even the smallest factors are to be given attention to pay careful attention to the peculiarities of the correlation coefficient we may be using. The product moment and its identities, e.g., the corrected rank formula, are the safest, but are not entirely safe. When the test scores for each variable are based on relatively few categories e.g., few pass-or-fail items, dichotomously scored rather than with many grades of success, the same sort of distortion occurs as with the phi coefficient but to a lesser and generally negligible extent. The tetrachoric, which is a computationally easy and therefore popular instrument, began by being roundly condemned by Hotelling (in connection with the principal components factorization) and then praised by Wherry and Gaylord (**136**) as a coefficient free from difficulty factor manifestations. According to Smith's more recent work, it produces difficulty factor artifacts to a lesser extent than the phi coefficient. The extreme values of the tetrachoric with strongly eccentric divisions *do* reach $+1$ and -1 but the approach values do not climb *toward* these extremes as early as they should, compared with the true, product-moment values. These difficulty factors are possible because the tetrachoric and other coefficients lose some of the information (e.g., by taking a cut on a normal distribution) present in the product moment, but they would not occur without eccentricity also. Error factors and the amount of variance lost in uniqueness might, however, be greater, because of the approximation involved when the assumptions involved in these coefficients are not met. The term assumption factors has been suggested for error factors of this origin; but none has yet been investigated. These questions await further discussion by statisticians, and though the correction suggested by Smith works empirically it needs more theoretical refinement.

More research is needed on the question of what coefficients of covariation can be used in factor analysis and with what consequences. While most studies have kept to safe coefficients requiring only slight corrections, if any, such as the tetrachoric, the rank order, phi, biserial *r, etc.*; we need information also about r_p (the coefficient of pattern

similarity), the contingency coefficient, (or other x^2 derivatives), the formula used by Thorndike (**122**), and a number of others with special convenience for special problems, such as Kendall's (**81**) tau coefficient ($T=[N(N-1)-4E]/N(N-1)$ where N is the number in the sample and E is the number of inversions of rank, i.e., individuals lower in one series than the other) used in Saunders' K-way scale analysis (**110**). An extreme instance of erecting factor analysis on a deviant coefficient is McQuitty's use (**95a**) of the common elements coefficient (**95**), replacing elements by people. This divides the number of people saying "Yes" to both items A and B by a function of the number saying "Yes" to each, whereby the correlation of A and B neglects any reference to people who say "No" to each. However poor the association of A and B this coefficient never becomes negative, and the meaning of "factorization" on such a basis—though it deserves to be explored—is so different in properties from the well-known techniques of factor analysis that it is probably inviting confusion to include it under the same name.

In general the best coefficient for continuous data is the ordinary Pearson-Bravais r. The substitution of the tetrachoric—a dichotomy on a continuum—has the slight risks of distortion indicated above and, because of its larger probable error, through loss of information, requires roughly a 50 percent larger population to get the same degree of certainty. "Phi divided by phi max," i.e., by the maximum possible phi for the given eccentricity of cut, is probably the best coefficient yet in use for dichotomized data.

UNITS AND SCALES

Coefficients cannot be adequately discussed before reference has been made to our next topic—measurement units and scales. Psychological measurement operates upon behavior and introspection. The former data can be expressed in three kinds of units (**17**)—raw or *interactive units*, *normative* units relative to a population scatter, and *ipsative* units relative to the individual's other performances. Introspective data can only be expressed in ipsative units, if any. Normative and ipsative units presuppose a continuum and are therefore applicable only to what have been called first-class scores, i.e., scores where various responses can be put in numerical order or rank on a continuum. Second-class scores, however, can also be used in some forms

of factor analysis. Here the absence of suitable units leaves us only with the fact that a response is greater than or less than, i.e., within or beyond, some arbitrary limit. There are thus only successive zones of response, usually two, e.g., yes or no, right or left, but perhaps three, such as conspicuously failing, getting by, and conspicuously succeeding. Third-class scores are those where the responses are put in two or more qualitatively different categories. Except where these can be brought into ranks or successive topological spaces they cannot be handled by ordinary factor analytic concepts, and the discussion which follows will therefore be restricted to first- and second-class score measures. In continuous, numerically expressible scales the chief discussions have centered on the effects of the form of distribution of the original data. As Thurstone points out (**126**), any distribution of original scores—rectangular, normal, bimodal—can be used, providing a normal distribution can be made of certain *derived* scores which are a monotonic function of the original scores. In other words, we can generally set out to rescale bimodally or oddly distributed scores to give a normal distribution to suit our convenience. For the original units of a psychological measuring instrument are rarely sacrosanct; they are generally arbitrary point scores on a test which is likely to vary in its difficulty in different parts.[2]

However, we need not even bother to transform the unevenly distributed scores. Neither the product moment r nor the principles of factor analysis assume or require a normal distribution. All that we lose by this omission is (a) a certain tidiness of the simple structure such as might have been obtained by normalizing and standardizing in the manner indicated above, and (b) the possibility of applying certain measures of significance that are very rarely applied. For, as Thurstone points out (**126**), the nature of the factors obtained (though not the particular variances and angles among factors) is remarkably immune to distorted distributions or crude coefficients. He asserts further that factor analysis can even be pursued with qualitative, all-or-nothing, noncontinuous scales and relations, but our opinion expressed above is that third-class scores do not yield factors falling within the usual concepts.

The one scale condition which destroys the effectiveness of fac-

[2] The exceptions are (a) absolute scaling scores and (b) what we have called interactive scores, when the physical units are quite definite and meaningful as they stand. There would not be much point in rescaling C.G.S. units, for example.

torization is that in which true curvilinear relations exist in the correlation plots. The transformation of raw scores which would correct the curvilinearity in one correlation surface would not necessarily do so in another, and when the curvilinearity is so complete that a score on one variable corresponds to two (or three) distinct scores on the other, no amount of rescaling can eliminate the difficulty. The correlation ratio η can then express the degree of relationship of the two variables; but at present there is no way of using η in factor analysis to give meaningful results.

The loss of a factor through normalizing *across* the columns of data has been discussed in connection with Q-technique in Part I.

DISTINGUISHING FIRST- AND SECOND-ORDER FACTORS

We must now pass on from the complications of factorization which spring from the mechanics of coefficients, or the nature of scales or the peculiarities of distributions—and which are either trivial and readily avoided or else absolute and inherent in the meaning of factorization—to a danger which is at once less trivial and less inherent in the mere mechanics. It concerns that risk of confusing first- and second-order factors which the student may have recognized in our earlier discussion of oblique factors and second-order factors. A second-order factor is a factor among factors, obtained from the correlation matrix of the factor vectors. If now we factorize a set of variables the vectors of which happen to coincide, in the common factor space, exactly with the directions of the primary factors in that realm; it is obvious that the factors we obtain at a *first* factorization will really correspond to second-order factors. But in the case instanced, we should not know it!

Such a coincidence of *all* variables with pure factors (plus specifics) is unlikely to happen, but it is more generally possible that *some* will. One must also consider consequences of the equally probable circumstance that the vectors will occupy roughly the same space as the factors even though they are not exactly aligned with them. The latter is likely to be approximated wherever variables are sparse and chosen from very disparate fields. For example, if we take an intelligence test, a measure of emotional stability, a physiological measure associated with surgency, a social index of dominance, and a speed ratio used for measuring schizothyme tendency; it is likely that we shall have taken

measures of primary factors A, B, C, E, and F, each of which, apart from a substantial specific, has little of any of the other common factors in it. Providing the factorization is accurate enough to see that this little is not overlooked, we shall not, in these circumstances, pass directly to second-order factors. And it is unlikely that a single variable will ever be so good a measure of a factor, unless and until it is specially so designed, as to have *nothing* of any other common factor in it. In addition, the almost inevitable presence of a specific factor in any single variable prevents it aligning itself with a factor to the extent that would be possible from the goodness of the reliability coefficient alone.

Nevertheless, the fact remains that until factorizations reach a high degree of accuracy with regard to sampling errors, communality estimates, etc., it will be a fairly common occurrence to have some areas of a matrix in which one inadvertently proceeds directly to second-order factors or factorizes variables which are for all practical purposes a mixed population of first-order variables and first-order factors.[3] This mixture of first- and second-order factors in the results of a factorization is a very real danger and is probably at the root of many misunderstandings, as it certainly was of the alleged incompatibility of Thurstone's primary abilities and Spearman's general-ability concept. It is necessary to develop experimental designs and concepts which can be more immune to these dangers of confusion.

With the present rarity of explorations of the second-order realm it would require us to go beyond experience to an a priori conclusion in order to offer a solution of this difficulty or to state how many orders are likely to lie one behind another in any typical, practical set of variables. Theoretically the number would certainly not be

[3] This difficulty may be illustrated by the recent study by Saunders and the present writer (**31**) in which one highly loaded variable was taken from each of about sixteen personality factors. Some of these factors were in behavior rating, some in questionnaire, and some in objective test media; and it was hoped that each real factor would be represented by two variables in virtue of the probability that one and the same real personality factor was represented by two or more media. Actually only about half the factors turned out to be the same personality trend in different media; and in the rest, one factor remained represented by only one variable. Among these latter the factorization seemed to proceed directly to second-order factors whereas where one of the primary factors was represented by two or three variables, it reproduced the primary factors again. It is this mixing of orders which is today an insufficiently realized danger with insufficient precautions taken against it.

restricted to two. By taking a very dense concentration of tests in some small area, one can generally multiply the number of factors found, just as, conversely, a sprinkling of very remote tests, no one of which measures more than a single primary factor, will be likely to lead at once to second or higher order factors. This must be considered also in connection with the splitting of factors discussed below.

SAMPLING OF VARIABLES

The solution must lie in developing more explicitly our notions about the density of a particular population of variables and in paying far more attention to sampling problems in regard to variables. For it is clear that with a high density of variables and a small matrix, we are likely to pick up factors of lower order than when variables are sampled from very diverse fields and brought into a single large matrix. Indeed, those researchers shortly to be mentioned who emphasize the importance of second-order factors are wont to say that some of first order are practically artifacts created by the narrow interests of certain investigators, i.e., that by multiplying very closely similar tests in a certain matrix there is a good chance of making a common factor out of what is really so particular a performance as to be best labeled a specific.

An attempt to provide an operational basis for an even sampling of variables has been presented by the present writer in the concept of the personality sphere (**22**). Possibly this can be developed more generally to provide a notion of the total area of possible variables and a standard density of variables or standard *scale* of operations in other areas than that of personality.

In the absence of such a theoretical development we can at least arrive at some concepts of relative scale by sufficient familiarity with the populations of variables in the actual researches so far published. We can file the factors in our memory, tagged with proper reference to the densities of the variables employed in those studies. For example, we know that the fluency of association factor (**22, 125**) has been found with a very diverse array of performance variables—anagrams, story completion, ink blots, creativity in drawing—whereas the various factors in auditory perception are found on tests mostly very similar in form. Though we cannot say that a given set of visual perception tests, or reaction-time tests, or reasoning tests are more

similar or densely packed in one battery than the other; yet we can recognize that a battery composed of one variable from each of these and from each of a number of other special researches is less dense, more spread out, and possibly going to give some factors which are of second order in relation to those found in the narrower batteries.

Though, in fact, people can agree in rough practical fashion as to the degree of density of a set of variables, it is desirable that the theoretical development sketched above should proceed. A need therefore exists in factor analysis for a means of determining what might be called the qualitative distance of one variable from another, a distance which is quite independent of correlation and based on qualitative differences or on some analysis of frequency of distribution in the environment. For example, it might be possible to take all the separate occupations listed in our culture and arrange them in chance (alphabetical?) order and take the ability performance required in each hundredth occupation in order to get a sphere of abilities with variables equally spaced over the realm of abilities.

Although there is no immediately adequate theoretical or practical basis for referring to the total population of variables of which any given battery is a random, a stratified, or a deliberately biased sample; yet effective users of factor analysis have come to realize that a proper technique for choosing a population of tests is quite as important as a proper technique in choosing the population of persons. Incidentally, this points to one of the present weaknesses which shows most clearly in Q-technique—that a random sample of tests is not so readily assured as a random sample of people. The nonsymmetrical nature of R- and Q-factorization shows most sharply here.

IMPORTANCE OF VARIABLE DENSITY

The proper sampling of variables, notably with regard to density, thus has as one of its aims the arranging of factors in tiers, having those of the same order in each, and also the avoidance of too much concentration upon factors in the lowest order, i.e., factors of a very narrow character. At present there is some trend among those most deeply engaged in research in this field to consider that second-order factors may turn out to be more stable and to correspond to more real and important scientific entities than the first-order factors. This opinion trend is based on little more than an intuition and may be

quite wrong, but its existence may be some corrective to those who unquestionably assume that the first-order factors are theoretically and practically the most important.

To be alert to all possibilities of confusion over second-order factors, the reader should note also the relation in which they stand to the variables on which the first-order factorization is based. A first-order factor may have its variance broken down alternatively into the variance of the variables (as when we take a column of the V matrix) or into the variance of second-order factors. This possibility of substitution of variables for second-order factors creates a real problem of interpretation. If I correlate measures relating to sun, moon, wind, etc. with the range of an ocean tide, and come out with a factor of tidal range, the measured attributes of the former appear only as variables in relation to this basic tide factor. I can say that the distance of the moon, etc. are loaded to a certain degree with this tidal factor. But by taking another set of variables—perhaps those dealing with consequences of tidal action—I might obtain tidal range as a first-order factor among them and later obtain the measures of moon, sun, etc. as second-order factors with regard to the first-order factor of tidal range and some other first-order factors in other realms also influenced by distance of the moon, etc. The latter could thus be on the one hand a variable and on the other, a second-order factor.

One may reply that this reciprocity of variables and second-order factors cannot easily be maintained because variables are (a) more numerous and (b) more highly correlated with the primary factor than second-order factors would be. (For the loadings in the specification equation must square to unity, whereas a column of V_n need not be so restricted, and generally far exceeds unity.) These differences may suffice in practice, and we should perhaps not concern ourselves unduly with this risk of confusion until the above speculative example has been replaced by a sufficiency of real second-order factorizations to indicate whether causality may run in one direction only with regard to the tiers of factor order.

At present, lacking any systematic basis in most fields for sampling variables, we can nevertheless take steps against the confusion of first- and second-order factors in the same battery. First we may take it as a good, rough, practical rule to design experiments with large

and varied rather than small and homogeneous batteries.[4] Every advantage—except the avoidance of work—lies in such a battery representing the main field of interest with sufficient density and then adding a fair number of variables chosen with great catholicity from various more remote fields. Such a procedure would reduce that prevalence in existing research of large specific factors (for absolutely unique factors are, on scientific grounds, almost certainly far more rare and smaller in variance than we now suppose. A wide variety of variables is naturally likely to give some higher correlations—and communalities—for any one variable). Also it would reduce the accumulation of orphaned factors, by which I mean factors appearing in one matrix that have no relation or continuity with factors from any other research, and are left uninterpreted.

FACTOR "FISSION" OR MULTIPLICATION

Among other problems created by insufficient attention to choosing the population of variables, particularly by failure to include landmarks from other researches, is that of identifying or matching factors from different experiments. The planning of continuity through landmark variables is discussed as a general technique in the next chapter and has been discussed earlier in factor matching, but may be considered here in relation to a problem of factor meaning, that involved in what is sometimes loosely called the splitting of factors. It may happen that after a factor has been known and labeled in a science for some time, another factor turns up much resembling it as to the loaded variables. For example, in psychology a certain factor was long known as verbal ability, but more intensive study seemed to show another factor, largely restricted to verbal performance, though with somewhat more emphasis on fluency, which also

[4] A *lower* limit to the number of variables to be used can be set if we know approximately the number of factors to be extracted. This is set by the fact that with too few variables the communalities which give the lowest rank to the matrix are not unique. With the same rank, i.e., number of factors, *different* factor patterns would then result. For defining this lower limit Thomson gives the formula:

$$r \text{ must equal or exceed } \frac{(2n+1)-\sqrt{8n+1}}{2} \qquad (49)$$

where n and r are the numbers of tests and factors respectively. When communalities have, as generally happens, to be estimated, this means we need at least 10 variables to make 6 factors determinate, 14 for 9, 18 for 12, and so on. But on general consideration one should decidedly exceed this lower limit.

could be called a verbal factor. Again, in the area of musical aptitude where a musical aptitude factor was first suspected, eight were later identified, loading the same or similar variables.

In such cases one is tempted to speak of fission, as if more refined research has split what was first regarded as a single factor. But this is not usually the correct way of describing such events. The additional factors are *new* factors not fractions of the old. They did not appear the first time because the variables which could share communality with one of the musical variables (in this example) were not present, and the musical variance unaccounted for by the first music factor was written off at the time as so much specific factor in each of the variables.

There is actually a sharp limit to such apparent multiplication of factors in a certain area. In the first place, if the researches deal with well-defined variables, as few as two may be enough to define the pattern of a factor. If only one factor has appeared in a battery containing these two variables, it is impossible for a second, new, factor to be found when they appear in some other battery, also accounting for part of their common variance. As stated above, a new factor in the general area, not discerned in the first research, can appear only when *one* of the variables in the first battery is linked in a new research with a wholly new set of associated variables which, however, may resemble in a general way those in the first battery. For example, a certain verbal comprehension test may have only one common factor (and a specific) in it when placed in a battery with a number of vocabulary tests and we may call this factor verbal ability. Then when it is placed in a battery with another set of tests of a generally verbal nature, but which happen to test largely fluency, a second apparent verbal factor appears, accounting for some of the specific factor variance left over in the first analysis. But this is really a fluency factor and has no role in any other of the first set of verbal tests except that *one* which happened to be shifted from one battery to the other.

It is, of course, possible to find a second new factor loading a set of variables in exactly the same pattern as the first factor—even when the variables are *precisely* the same in the two distinct factors. But in this case the two factors are bound to appear from the beginning in one and the same factorization (if there is enough hyperplane stuff to give distinct hyperplanes), so there is no question of whether

a *later* splitting of a factor is called for. However, there is also a definite limit to this simultaneous appearance of factors repeating one and the same pattern. This limit can be most simply illustrated by considering a pattern in terms of two variables only. Now the squares of the loadings of factors affecting any one of these variables cannot exceed unity. For example, two factors in this area cannot both have loadings above 0.70 each in one of these variables, and three cannot simultaneously exceed 0.57 each. If a loading is not much above 0.5, it is not possible to consider the factor as being highly characterized by that particular variable. For example, if a factor had its *highest* loadings of 0.5 in a couple of music tests, we might well hesitate to call it a music factor; for it would be reasonable to expect that much higher loadings would be found for it by exploration of some other field. It might for example be essentially a general auditory memory factor of *some* value in the musical field.

There is, thus, in precise experiment no risk that views as to the major factors operative in a certain area will need constantly to be changed in response to each new research. A spatial ability factor with loadings of around 0.8 in each of three characteristic spatial perception tests can be depended upon to be reproduced with invariance in other experiments, and does not leave much room for a second spatial ability factor in any significant use of the term. But experiment is *not* always precise or adequately planned, so that in the actual history of research, it happens that doubt sometimes arises as to the number of factors in a certain area and as to the identity or separate existence of two factors. Experiment fails sometimes because a factor occurs in only two or at most three variables in the first research and though these may be theoretically enough they are not always in practice enough to define the loading pattern; or it fails through confusing first- and second-order factors, or through poorness in essential techniques, notably insufficient factor extraction and rotation, so that coöperative factors (page 285) have their high variables treated as a hyperplane and have their variance treated as that of a single factor. (This somewhat condensed description can be expanded by references to pages 528, 531, and especially to Diagrams 19 and 20, of 22.)

The best proof that there are two distinct, powerful factors in an area is to obtain them simultaneously in a single matrix, instead of trying to show that two distinct researches, not having identical

variables, deal with similar but distinct factors. This can best be done by attention to choice of variables additional to those which have first shown the two factors and which will (a) be numerous enough to define both hyperplanes and (b) help bring out the communality while being very differently loaded in the two factors. As pointed out in dealing with coöperative factors and again in the above discussion, there is no theoretical impossibility in the notion of two distinct factors having exactly the same loading pattern (providing no variable exceeds 0.7) and there is some indication that such similar patterns occur with greater frequency than one would expect from chance.

DEFINING THE NATURE OF A PARTICULAR FACTOR

This problem of encountering similar factors in successive researches, and which we see is better called one of factor multiplication than one of factor fission, brings us naturally to the last and scientifically ultimate problem in the present discussion of factor resolution and fixation, namely that of *defining* any given factor. Like anything else a factor can be defined denotatively and connotatively; that is to say, we can point to several examples of its action or we can assign attributes to its essential nature. The discovery of a factor in several contexts, especially its discovery by all three techniques (P-, T-, and R-), and in matrices with wide ranges of diverse variables (additional to those few landmark variables which have to repeat themselves to fix it) assists the extensiveness of the denotative definition. The meaning of the factor is now fixed because one knows definitely where to find it and in what company.

The connotative definition—the attempt to give the essential nature in general terms—can be approached by a hypothesis (for an ultimate definition has first to serve its apprenticeship as a hypothesis) working either from below or from above. Working from below means building up a concept as an empirical construct, while the approach from above involves importing a logical construct from some remoter field of reasoning. For example, one might observe that all the variables highly loaded in a certain factor in sociological data involve aggressive social responses, and call it a factor of social aggressiveness. On the other hand, a person might harbor a theory that one of the most powerful factors in this region is economic deprivation, and he may choose to interpret all these loaded variables as consequences

of economic deprivation although he cannot see a single sign of economic deprivation in the empirically given variables. In the writer's opinion, to begin with a logical construct is as inept here as it is anywhere else in the early stages of scientific investigation. The factor should reach at least a first level of definition as an *empirical construct* before it is profitable to venture upon that reasoning by analogy and that departure from operationalism required in logical constructs.

The definition of a factor as an empirical construct follows remarkably closely the procedures stated by Bacon and refined by Mill for arriving at the essential nature of anything. One observes where it is conspicuously present as a positive influence, where it is conspicuously present as a negative influence, and where it is generally absent. For example, the surgency factor in personality has conspicuous positive loadings in the traits cheerful, happy-go-lucky, in proneness to errors in certain tests, and in high electric skin resistance, while it has conspicuous negative loadings in anxiety, depression, seclusiveness, and slow, exact performance. It has no relation to various intelligence and ability measures, to emotional stability, degree of dominance, level of education, etc. These presences and absences suffice to give a reasonably clear picure—an empirical construct—of the dimension with which we are concerned, and lead up to the more theoretical construct of a physiological proneness to acquire inhibitions which operates more in the desurgent than in the surgent individual.

In interpreting and defining a factor, it is essential also to keep a constant reference to other *factors*; for these exclude, by already representing, certain interpretations one might otherwise be tempted to adopt. Most factor interpreters naturally tend to place emphasis on inspection of those variables wherein the factor is conspicuously present and here the procedure of abstraction from these variables is the same as in all concept formation and all invention of universals. Like a person struggling with a classification test, the researcher gropes for some characteristic common to the selected variables. This groping is likely to become better directed and more systematized as we come to know more about what indications will illuminate the *general* nature of the factor, e.g., whether it is likely to be merely a dimension or a substantial entity, etc. For at present, in psychology for example, we may need to look for a common introspectable mental

process, or an unconscious reaction pattern, or a similarity in the actual test content (e.g., information about biology), or a similarity in test form, e.g., eduction of correlates, and so on. In sociological patterns the possible explanatory bases for the common fate of the variables are still more varied, including historical accidents.

Whatever hypothesis at the empirical construct level may be yielded by an intelligent study of comparisons and contrasts and the resulting attempts to educe relations,[5] the next step before the scientist is that of testing this hypothesis by repeating the experiment with a substantial block of the variables, unchanged except for the new crucial variables being added. Of the latter some will be designed to express the quintessence of the factor as now conceived (positively or negatively) and some will be such that on this hypothesis they should *not* show the factor at all.

If this first hunch is in the right direction, a still clearer picture of the factor will now emerge because of the higher loadings of the crucial variables achieved in the second experiment. In practice it seems that two or three rounds of experiment are likely to be necessary before one can hope to pass from a concept closely tied with the actual tests to one with the character of a logical construct, but an inspired guess may hit the bull's-eye sooner. In a personality rating factorization, the writer obtained one pattern among variables which suggested the influence of general mental capacity and he entered upon the hypothesis that the pattern was the effect of intelligence upon personality. The introduction of an actual intelligence test yielded a pattern of correlations with these variables exactly similar to their loadings in the factor. The same mode of identification, in this case of the factor of general neuroticism, has been successfully used by Eysenck. In neither case, however, did the pure measure actually give a perfect saturation in the factor. The fact that perfect saturation could be obtained by various corrections (notably for attenuation) is not in itself convincing, since we know that surprising agreements can sometimes be obtained with unidirectional corrections. Checking a hypothesis by demonstrating an identification could be better carried out, as explained in Chapter 17, by correlating the two uncorrected loading patterns or by other methods there suggested.

[5] This eduction may turn out to be most successfully performed as an unconscious or intuitive activity. One should soak oneself for some days in the evidence as to the nature of a factor and then sleep on it.

In the end the identification of a factor cannot be considered complete merely through achieving such an agreement of pattern or obtaining perfect saturation of the crucial variable introduced to test the hypothesis. It is necessary to check by controlled experiment, rendered possible by the identification. For example, if a pattern to all intents and purposes is fixed as that of hyperthyroidism, the proof is completed when physiological experiments show the variables in question to be responsive to the agent identified.

FACTORS DEPEND ON VARIATION

While dealing with factor resolution and the meaning of factors, it is necessary to emphasize one general point implicit in the statistical expressions. The description of the specification equation pointed out that where a certain performance is due to two or more factors, it would be possible for two people to achieve the same score without being identical in their factor endowments. That is to say, the same performance is accomplished by different combinations of influences. For example, of two equally good tennis players, one may have more intelligence and another, more agility. Perceptually these differences of combination are referred to as differences of individual style. One must not overlook the possible extreme of this argument—that a person, or indeed a substantial block of the population *might have no endowment whatever in certain factors* demonstrated in factor analysis. (In P-technique the person would be bound to have endowment in any factor which appeared, but there could be whole series of *occasions* on which he did not show it.) Normally, of course, we should find factor endowments to be normally distributed, but factorization can also be carried out with all-or-nothing distribution using biserial r's. In general, therefore, R-technique tells us only what is typical of the population; it does not tell us that each and every individual performs in the given variable by using all the factors indicated in the specification equation.

It is perhaps appropriate to remind the reader, while pondering on factor meaning, of something also mentioned earlier, namely, that the factor loading (situational index) is not a measure of the *mean amount* of the contribution of the factor to the situation. For example, the discovery that in a certain collection of books, the factor of weight is loaded 0.6 in thickness and only 0.2 in height simply indicates that for a given weight (overall size) these books vary

more in thickness than they do in height—as books on a tidy shelf should. By contrast with the specification equation, we have to consider two other sets of relationships, namely (1) that books are absolutely taller than they are wide, and (2) that in computing the volume (and weight) of an individual book all dimensions are equally important. Or again, where the relation is strictly additive, we may find that the mean contribution of a variable to the factor is quite different from its loading. For example, variation in the height of men might conceivably be associated more strongly with variation in neck length than leg length, but the legs contribute more absolutely to the height.

Doubtless, due to the variation of parts tending to have a fairly constant relation to their absolute size, the mean absolute contribution and the loading will *in general* be of the same order. But the investigation of a realm of data will require first a factor analysis to indicate what variables are involved in a factor, and then, for some purposes, a further, nonfactorial research to see *how much they are involved absolutely.*

Questions and Exercises

1. What are some of the kinds of quantities that discovered factors may represent? Describe the relation of efficacy to degrees of functional unity.
2. Describe at least two circumstances which lead to spurious factors.
3. What measures of covariation have been and can be used to obtain matrices for factorization? Discuss phenomena of eccentricity factors and the ways in which they can be avoided.
4. Discuss the effects of (a) departure from normal distribution and (b) variations of mean standard deviation in the battery as a whole upon correlations and the results of factorization.
5. How do possibilities of confusion of first- and second-order factors arise? What possible bases exist for speaking of density of variables in a given research?
6. Indicate what are the lower limits n to the number of variables in a factorization anticipating the need to make determinate a certain number of factors r. Why is it desirable to work with large matrices and to pay special attention to selection of a suitable population of variables?
7. Discuss the limits to the multiplication of factors in a given realm of variables. By what steps are individual factors best fixed and interpreted?
8. In what scientific explanations must we beware of misunderstanding the true meaning of the situational indices?

CHAPTER 19

The Chief Manipulatable Features in Classical Factor Analytic Experiment

Resolution, fixation, and interpretation of factors involves, as the preceding chapter shows, quite as much attention to the manipulation of initial design of experiment as to the subsequent processes. Our concern there was mainly in manipulating design for a few specific ends, however, and we may now aptly proceed to a *systematic and comprehensive* survey of what proper attention to design can do for the scientific usage of factor analysis. Once again, as in the preceding chapter, it is understood that we discuss the question in regard to factorial methods useful to the scientist, i.e., within the framework of multi-common-factor constellations and rotation to simple structure or some other inherent natural structure.

CLASSIFICATION OF CONTROLLABLE CIRCUMSTANCES

With this proviso as to adequate statistical methods of analysis, the experimenter is free to manipulate in the design of his factor analytic research the following conditions:

1. The variables (responses, performances, attributes within given stimulus conditions)
   a. As to their nature
      i. Kind, choice of species
      ii. Sampling selection, as to scatter and level of eccentricity, e.g., difficulty in relation to the population *within* the species
   b. As to their number
2. The organisms or units
   a. As to their nature
      i. Kind, choice of species
      ii. Sampling, scatter, and level
   b. As to their number

3. Methods of scoring and correlating the measurements
4. Use of P-, T-, or R-techniques and their transposes
5. The specific relations of any of the above to some other experiment or research, since in parallel profiles and in many other as yet unexplored factor analytic uses the essence of the design consists of two or more factor analyses arranged strategically in relation to one another.
6. The environmental and time (including controlled experimental) conditions in which the variables are measured. Some manipulations here come near to taking the design out of the classical, constant situation design into the varying stimulus factorization described in the next chapter.

Let us consider the questions on the above scheme briefly, in order, before turning to a more expanded treatment of those which need it. It will be noticed, as mentioned when discussing Q-technique, that the manipulatable aspects of (1) variables and (2) organisms above are mutually symmetrical except that (1) kind or species in tests is not quite equivalent to kind or species for organisms. Organisms tend to fall into well-defined species, i.e., classes of essentially similar individuals from which a sample is taken, whereas tests or responses do not have the same definite, naturally-produced boundaries to their groupings and (2) as a correlative, test selection is not the same as organism sampling, due to our inability adequately to define the parent population from which the tests are supposed to be selected. These remarks apply to current practice. If we like to break with convention and usefulness and factorize a population of organisms not all of the same species, then we are truly operating with a situation symmetrical to that for tests. In the usual practices, symmetry also fails, as pointed out in Chapter 7, through one test having the same average as another (because of standard score alone being meaningful), whereas an individual may be above average in all tests.

As to question 3 above, different methods of scoring and different coefficients by which to express the relationships have already been discussed sufficiently for the present volume in Chapter 18.

On question 4 the decision as to the use of P-, T-, or R-technique interlocks with the choices made in 1 and 2 above. If there are many persons, we may be tempted to choose R-technique; if few, then Q-technique, provided that many tests are available. Nevertheless, as indicated in Chapter 7, there is more to be considered than these matters, and the decision for R- or Q-technique, P- or O-technique, etc., is best made in its own right, not merely as an expediency after the conditions of 1 and 2 have become settled. The decision in section

4 as to using P-technique also interlocks with the decisions made in sections 1 and 2, but again the choice of the technique in relation to scientific purposes is the primary consideration.

The conclusions already discussed on the relative utility of these methods may be summarized by saying that somewhat greater difficulties beset Q-technique by reason of the missing first factor, the absence of a criterion for rotation, and the uncertainties of interpretation. But one must heed the relative accessibility of the data required for each (number of subjects, time for measuring, duration required for longitudinal studies) and the predictive purposes immediately in view. As to P-technique, it presents an intrinsic difference in the meaning of the factors obtained, but can clear up issues where R- or Q-technique would fail and can be used with great strategic advantage in company with R-technique. For some factors may be expected to reveal themselves more readily by P- than by R-technique by reason of greater variance in the longitudinal situation. However, if the object of search is to determine the form of common factors, P-technique, whatever its other advantages or conveniences, cannot be recommended on grounds of economy; for whatever testing time is gained in factorization of the single person is lost by the fact that factorizations of other single cases must be made before one can be certain that the pattern obtained in a single study does not have absolute uniqueness but only the relative uniqueness of scattering about some average shape. The essential comment on these methods is that at some time all six should be employed on the same problem, for until this is done our understanding of the relative efficacy of the factors and of their behavior in diverse circumstances makes our definition of them incomplete.

DESIRABLE CONDITIONS IN CHOICE OF VARIABLES

With this view over the relations of the manipulatable conditions we can now turn to a more intensive examination of each. Beginning with the nature and number of variables we find that the major aspects have already been dealt with in the last chapter, and various minor aspects have been encountered throughout the book, so it remains only to summarize our conclusions in a single list as follows:

1. Unless variables having variance in a certain factor are introduced, the analysis cannot reveal that factor. This does not mean that you only get out what you put in because (a) the factors which

emerge may, and generally do, turn out to be different in nature and number from those which the experimenter *thinks* he is putting into the battery. His choice does not settle the outcome. (b) The experimenter merely puts in *variables*, but the analysis gives him factors and a structure among them. To say that this structure is only what he knew beforehand is to say that the cup is nothing more than the clay from which it was made or that there is not more in heaven and earth than is dreamt of in our philosophies.

2. Although the experimenter should insure proper density of variables manifesting the phenomena with which he is principally concerned, he should make a special point of (a) balancing this with a sufficiency of representatives from other areas and (b) attending to proper density in the sample as a whole. The first is necessary to provide hyperplane stuff and to give orientation to known factors.

3. As to density it is necessary to have in mind the concept, however rough, of the total population of phenomena from which it is taken. This is necessary in order to get an even density of representation that will not give mixed first- and second-order factors, that will give the position of any factors of particular interest in relation to most factors in the field, and that will provide a sufficiency of hyperplane material for whatever number of dimensions may prove to be involved.

4. For the last reason and also to give clear distinctions when factors multiply in a certain field, a large matrix of variables is vastly preferable to a small one and a certain lower limit in number of variables must in any case be exceeded if the rank and factor structure of the matrix is to be determinable with inserted communalities.

5. Marker variables should always be included from the high and low points of factors in previous related research—two high markers being essential, and three preferable, for each factor.

6. It is necessary to remember that the precise definition of the nature of a variable, instrumental in comparisons of factors with other research factors, requires not only the usual exact definition of administration conditions but also a statement about the population reacting to the test or the situation.

7. Advance in understanding the meaning of factors depends upon a cycle of hypothesis formation (empirical constructs) followed by test design (or variable choice) guided by these interim hypotheses. Generally the task is to choose or invent variables for subsequent fac-

torizations which contain to a higher degree (and ultimately to the point of perfect saturation) the factor essence. This is approached by reasoning as to the universal (in the sense of logic) character which can be abstracted by educing relation among features of the variables initially shown to be highly loaded in a factor.

8. In factorization work the variables employed need not all be on the same level in the sense of being, say, all responses to a situation. This will be more clear when we have discussed in Chapter 20 the possibilities of freer experimental designs than have traditionally been used, but may be illustrated here by the notion of including the criterion (of success, when trying tests against an occupational success criterion for example) in the matrix with the variables. Some of these changes may lead to entirely new hybrids of factorization and experiment. More generally, however, this broadening of the nature of variables merely means inclusion of environmental conditions and inherent characteristics of the subject, e.g., age, along with regular test scores.

Turning now to the second aspect of variable selection—in regard to difficulty level and variability as classified under 1. a. ii. above—we perceive that in its usual sense this is a matter of relation between the variable and the population, and so is best postponed for discussion under 2. a. ii. In the sense in which variable selection is symmetrical to population selection this question concerns itself with the existence of uneven difficulty among the various tests, and here few practical problems arise. As the discussions in the previous chapter indicate, all considerations emphasize the desirability of using tests as nearly as possible equal in eccentricity. Since the factor patterns of even genuine factors are going to appear distorted if the reliabilities of variables differ considerably, it is desirable that variables be of equal difficulty and variance, for on these, among other things, the reliabilities will depend.

DESIRABLE CONDITIONS IN CHOICE OF POPULATION

The important issues of designing the population—item 2. a. i. above—should properly begin with an almost philosophical question which has so far been neglected, namely, what species of organisms or, indeed, of entities may be allowed to make up our population? So long as we deal with people, or rats, or social groups, the cost of articles, etc.; the entities are so taken for granted that we are content to ask

only about homogeneity in the sense of avoiding bimodality of distribution, etc. But more recent developments begin to raise questions as to what the meaning of a correlation, or a factor, becomes when we employ more unusual units of entry. In P-technique and certain factorizations in economics, the population may consist of years. Small groups (**25**), whole nations (**27**), cities (**122**), city districts (Tryon's work), occupations (**20**), etc. have constituted the species entered into correlations in sociological analysis, and so on. Chemical substances (**66**), points on the earth's surface, actions of bacteria, and even Cheshire cheeses (**2**) have formed the population elsewhere. Is there any limit to the sorts of things on which we may hang our measurements for the initial correlations in factorization, and in what way will the meaning of the factors differ according to the degree to which these things have individual organic unity? Let us remind ourselves for proper perspective here that in the covariation chart (page 109) the individual measurements have their locus defined in three ways: (1) by a moment in time at which they are made, i.e., an occasion with its accompanying stimulus conditions; (2) by the organism from which they originate—there called a person; and (3) by a response in a defined set of measuring conditions, there called a test. To define the whole situation, we should also have to fix (4) the observers who make the measurement—principally whether objective, behavioral data or measurement by the introspection of one uncheckable observer; and (5) the scale units in which the observations are made. A fuller discussion of these five aspects of a completely defined response measurement has indicated that for most purposes it suffices to concentrate on the first three (**151**).

As we have seen, these organism-occasion-performance dimensions placed in the covariation chart can be employed in alternative correlation series to give R-, Q-, and P-technique solutions and several others besides. R-, Q-, and P-techniques throw light on the structure of organisms or an organism, but O- and T-techniques can instead throw light on the organization of occasions. Thus, if the correlated variables in the matrix consist of, say, fifty occasions in the individual's life history, and the series consist of his reactions to a lot of interest-attitude situations, the factors will be historical phases of his personality in any one of which his reactions were consistent, e.g., moods, maturation stages, or multiple personality phases.

However, in all the changes that can be rung on the dimensions of the covariation chart, e.g., of absolute scores in R-, P-, and Q-techniques, increment scores on R-, P-, and Q-techniques, the correlation of occasions, etc. (**22**), one thing has so far been assumed to be constant. The population of *beings*, i.e., the units with respect to which the population of performances (tests) and population of occasions has meaning, is normally a set of organisms. It is a collection of persons, animals, small groups acting as groups, culture patterns or some such unities of the highest order of unity, i.e., that unity which is indivisible.

It is at present only of relatively speculative and philosophical interest to inquire in what way the meaning of results would change if we took measurements on things other than organisms as our reference points. For example, we might, as in the few eccentric studies mentioned above, correlate characteristics for points on the earth's surface or the attributes of a number of chemical substances or physical states. In general it would seem that what we find by factorization must be considered a characteristic of the particular degree of organization we take. If we take temperature and humidity, etc. as variables with respect to points on the earth's surface, we end with generalizations which apply to any point on the earth's surface in virtue of the general laws of weather on the earth's surface. If, on the other hand, we deal with variables with respect to entirely discrete organisms, our factors have to do with the typical internal structure of these organisms.

Most actual research situations are intermediate, and consequently the factors deal both with the structure of the organisms and with effects produced by the field in which the organisms exist. For, in general, the organisms we deal with are *not* entirely independent; they create a field for one another or else are together embedded in some larger organism. This point first attracted the attention of social scientists in some debate as to the legitimacy of Thorndike's (**122**) correlation of characteristics of cities (as contrasted, for example with using persons, or whole nations). It was pointed out that since cities in one country are not politically independent but are participators in the same culture matrix, their variability is constrained. The answer would seem to be that our independent persons are in fact never completely independent events. They interact and live in the same culture. The factorization therefore expresses not only the internal structure of the organism (due to heredity and social influences) but also the effect of the present pressures and provocations,

the covariations produced by the form of the whole culture. The latter are perhaps especially revealed in the second-order factors. For example, the correlation of certain psychological defects (e.g., low intelligence, low degree of respect for law) with local economic defects might be greater with nations than cities because nation-wide relief measures wipe out some of the intercity variations and the causal connections within the pervasive national organization. Or again the correlations and factors obtained between, say, emotional stability and sociability respectively for individuals and families as correlation points might differ. For the unpopularity of the emotionally unstable might operate more powerfully within the family, i.e., within an enduring group of individuals, thus producing an association of unsociability with instability.

This leads us to two conclusions not previously made explicit. First, the factor patterns found from the same sets of variables will differ according to the organisms or aggregates taken as units of entry in the correlations. Secondly, in the correlation, the factor does not necessarily belong wholly to the organism or exist within it as a property distinct from any property of the environment. Hsü (78), for example, objects that a diurnal fatigue factor found (23) by P-technique is a function of the hour of the day rather than a property of the organism. But even in the classical factor analytic design (as contrasted with the hybrids shortly to be described) in which there is no experimental manipulation of the environment, the factors found are just as likely to represent structure in the organisms environment —in so far as it is held constant—as they are to represent structure intrinsic to the organism.

Thirdly, it is necessary explicitly to point out that there is a second reason for differences in factor structure found between situations where the organism is an individual point and where it is a group point. This arises no longer from differences in the degree of organic independence of the points in these two universes, but from sheer mechanical, statistical effects. For example, the correlation between intelligence and size of family, if we take families as points, is lower than if we take city districts as points. In general, any grouping in which the means of the groups replace individual scores gives higher correlations than would be obtained by using individuals, because chance error and some other sources of irrelevant variation are much reduced by the grouping. Indeed, the correlation will get larger with increasing size of groups used, according to definite statistical law.

SAMPLE SELECTION EFFECTS

With this brief discussion of problem 2. a. i. (page 342) concerning the nature of population members, we can now turn to 2. a. ii. concerning sampling, scatter, and level. So far our main conclusions have been that sampling selection has the following effects:

1. When it increases the variance of the population on all variables used, it is likely to produce higher reliabilities and therefore a greater number of significant factors.

2. That a population selected for a high mean *level* on all tests, i.e., for easy tests, or a low mean level, will normally alter the variance and intercorrelations of factors but not the number of factors or their recognizable loading patterns. This is part of the general problem of univariate and multivariate selection, the extended theoretical treatment of which is dealt with elsewhere (**7, 120, 126**). However, though changes of mean and sigma will not theoretically alter the loading patterns, etc., it remains true that the clearest factor structure is likely to be obtained using a middling level of difficulty on all tests and a substantial sigma.

3. When the population is selected in regard to tests (or tests in regard to population) so that some are very difficult and some very easy (i.e., high eccentricity) and when the measures are of the kind which break down into discrete items few enough to offer relatively few pass or fail responses per variable, spurious factors will be created which are largely due to the failure of some indices of covariation to avail themselves of all possible information. Thus when we say 60% of those who pass on variable A also pass on variable B, it is uncertain (with a simple pass or fail grade) how this 60% is taken from the various levels of those who pass. Through this weakness, the phi coefficient and to a lesser extent the tetrachoric, are incapable of giving a proper proportion of high correlations, and the former cannot even give a perfect positive or a perfect negative correlation (1.0 or -1.0) when perfect correlation exists. A possible correction for this is suggested which requires further theoretical examination, but which has successfully removed spurious eccentricity factors in some actual examples.

4. Although reduction of variance of tests produces only relative effects, reduction to the point where no variance at all exists with respect to a given factor would naturally remove the factor. This can

be achieved (so long as tests have any variance) only by *selection of the population* with respect to one *factor*. For example, a college population is sometimes selected to such a narrow sigma for intelligence that no significant intelligence factor appears.

There are thus two reasons for the failure of a factor to appear—the absence of tests from the battery that have to do with such a factor as discussed above, or the absence of persons who deviate in any degree from the mean with respect to the variables involved. It is thus a fundamental limitation of the factor analytic method that it cannot reveal any functional unity, no matter how real, *if everyone possesses it to exactly the same degree.* (In the case of P-technique the equivalent condition is that the individual never varies in it.) Thus if we R-technique factored all automobiles of the same make, we should *not* get a factor of body length though all automobiles have a length dimension, and if we P-technique factored the performances of an airplane, we should obtain no factor for the power of its engine, providing that power remained unchanging. However, it has been well said that this is a limitation to almost any experimental or investigatory method. If the variables cannot be made to vary, little can be found out about them. The difference from a research standpoint, however, is that this is more likely to happen without our knowing it in factor analytic work than in controlled experimental situations.

Fortunately in the organic performances of the biological and social world such adamantive invariance rarely occurs. Nevertheless it is important to bear in mind that (a) it is desirable to attack every problem by both R- and P-techniques in the expectation that factors which have too little fluctuation to reveal themselves clearly by one will be caught by the other; (b) it may be necessary, if one suspects that some factors are being missed, experimentally to produce greater variance by various extreme selections or provocations; (c) the importance of a factor is not to be judged by the magnitude of its variance, for quite basic dimensions and powerful influences may happen to fluctuate little; and (d) occasionally, despite all that can be done under (a) and (b) above, it can happen that factor analysis misses a factor in a situation.

This discussion of population variance and selection may help to bring out more clearly the difference of loading value and absolute contribution value discussed at the end of the last chapter. From the statements there made, it may be obvious to the mathematically minded

that the situational indexes in the specification equation express the extent to which *variations* in factor strength affect variations in performance. But in spite of that warning, many students continue to fall into the habit of thinking that these loadings represent the *absolute mean contribution* of the factor. If a man gets his income partly from a grocery store and partly from gambling on horse races, it may be that the fluctuations in his total income are much more closely tied up with fluctuations in the latter than in the former, but the former is likely nevertheless to average the larger contribution. As stated, since there seems to be some tendency for the coefficient of variation to be roughly the same for different variables, there is *likely* to be some parallelism of loading and mean total contribution. But in general the aim of factor analysis is to *determine what influences are at work* and to leave to other, generally more elementary methods, the determination of the mean absolute contribution of these influences and in fact, of any other values beyond the loadings themselves.

5. Discussion of the effects of manipulation of variance of the population either by selection or by bringing experimental provocations to bear raises the question as to whether there is an *ideal* population to be aimed at in respect to mean and sigma. Most discussion has centered on the latter, and particularly on what degree of homogeneity of population is in general desirable. Spearman, intent on defining one factor—general mental capacity—not only threw out variables which introduced other factors but also threw out populations which would do so i.e., he advocated homogeneity of population with respect to age, sex, education, etc. For, as he pointed out, if the population varies in age, increases in intelligence would tend to be correlated with more years of education, physical growth, etc. so that the factorial definition of intelligence would come to include not only ability but also experience, physical stature, and even the number of erupted teeth!

HOMOGENEITY OR HETEROGENEITY

While admitting that Spearman's purification of the general factor of ability (**113**) by this control of population was highly desirable, we may yet question whether it actually leads to the acceptance of the general principles of design to which it appears to lead. If Spearman had been using multifactor analysis, and if a proper catholicity of variables had been introduced, these extraneous influences of ex-

perience, physical growth, etc. would themselves separate out as factors distinct from the general ability factor. As such distinct factors they might *correlate positively* with the general ability factor, because through the childhood range physical size *does* correlate with intelligence; but they would no longer be mistakenly viewed as part of the general ability factor itself.

Some reduction of population heterogeneity is desirable when one wishes to rule out associations that are accidental. But accidental is a word for statisticians to toy with! No correlation, or resulting factor loading, is accidental if it sufficiently transcends the limits of chance. If we feel that such a relationship is accidental, it means that we have still to seek for causes beyond our present understanding. But there may well be good reason, nevertheless, for our belief that one variable is less intrinsic to a factor than is another, though they have the same loading in it. For example, the size of a car is likely to correlate both with its horsepower and with the wealth of its owner, but we feel that the latter is not so intrinsic a part of the description of a car as is the former. It seems likely that most of such instances of accidental association are properly instances of substantial correlation between two distinct factors (car size and social status of owner in the above case). Their correlation must in turn be explained by some second or higher order factor influencing both. If only one variable representative of the second factor is included, its positive correlation seems to throw it in the first factor; but if more variables from the second were included, it would stand out as a second factor.[1]

The problem we are now considering is sometimes approached in another way without actually manipulating the population variance in the experimental design, namely, by partialing out (by the use of partial correlation coefficients) the influence due to a certain variable before factorizing the coefficients. Substantially the same pros and

[1] As a hypothetical example to show more clearly this effect of more or fewer variables in a heterogeneous population, let us take a simple physical example. If we took measures of height, shoulder breadth, weight, etc. in a mixed Japanese-Negro population and included an index of skin nigrescence, it is likely that darkness of skin would appear by its loading to be part of the general size factor; whereas in fact we should regard it as accidental and extrinsic. But if other variables were added, e.g., measures of lip thickness and Mongolian eye form, we might expect two factors, corresponding to the racial patterns, to separate out. Part of the size variance would also go with these, but presumably a general size factor would remain, purified of variance in these extraneous variables, distinct from the racial pattern factors but correlating with one of them.

cons apply here. Sometimes, however, the objective seems to be to partial out what would otherwise appear as a single factor. For example, it is not unusual to partial out the trend from time series of socio-economic data. This is unnecessary, for a properly conducted factorization would probably itself partial out the trend as a distinct factor or factors highly loaded with the time variable. Moreover, this latter partialing out has the advantage that it sets aside from the other influences in which one is interested true factors, whereas the prior partialing out of variables by the experimenter is likely to take out all sorts of mixtures of factors.

One disadvantage of permitting full, normal heterogeneity to exist in the experimental population is that thereby the variance of the individual factors, because they are now more numerous, gets cut down; so that the perception, fixing, and interpretation of the factor loading patterns is more difficult. For example, in the example of Spearman's research above, a normal heterogeneity would allow at least some of the variance in neogenetic test variables, e.g., analogies, classifications, to go into age and education factors, while some of the variance in a verbal ability test might disappear from the general ability factor and appear in a sex factor which was not possible in a sex-homogeneous population.

This limitation to clarity of individual factors in normal heterogeneity has also led, curiously enough, to an argument for experimental design which at first sight seems to run directly counter to Spearman's. This is Thurstone's prescription for design in which (**126,** pp. xii, 440) he argues, "Select the subjects so that their attributes are as diverse as possible in the domain to be studied." This agrees with the general biosocial-scientific principle of studying exaggerated and abnormal forms in order to magnify what is present in the normal. Actually the Spearman argument for homogeneity and the Thurstone argument for heterogeneity are not truly opposed but have the same objective—namely, the increase of variance of the factors in which one is interested. Thus, in the case of general ability, Spearman's request for homogeneity was for homogeneity in respect to things *other than intelligence*—he would gladly add a sprinkling of extreme instances of ability, as in defectives and geniuses—while Thurstone's is a claim for positively increasing heterogeneity *in the thing studied*.

Unfortunately there are dangers and limitations to either of these ways of magnifying factors. One does not know beforehand, in ex-

plorations of new areas, what factors exist to be magnified i.e., what is the thing studied. Relatively blind exaggeration of heterogeneity on the variables which *seem* to contain what we want may actually obscure the factor structure and produce an excess of those accidental associations just discussed. (Compare comments on "Criterion Rotation," page 250.)

This can be realized more acutely if we take instances of homogeneity which rule out certain features which naturally go with the variance of the factor. For example, we might take variation of physical performances with age in a group of *perfectly healthy* people scattered over a wide age range—ignoring the fact that some increase of physical ills is at least a statistically normal accompaniment of age. Or, when increasing heterogeneity we may do so to the extent of incorporating two distinct species, or, at least, a bimodal distribution in the matters which most interest us. In this connection we may take the illustration of Eysenck's being rather severely criticized in his study of the neurotic for allegedly making his neurotic population out of a group of conversion hysterics and anxiety hysterics. It is argued that his second factor (after the extraction of the general neuroticism factor) is a conversion-vs.-anxiety bipolarity simply because the population had this two-species character. Whatever may be said about the absence of simple structure and the failure to extract further factors in this study, it cannot be argued that this factor (which corresponds to surgency-desurgency) is an artificial product of the population selection for heterogeneity. Rather we must regard the absence of *other* personality factors as the price paid for the magnification of this particular factor by choosing high heterogeneity.

In conclusion, the designing of populations to be relatively homogeneous or heterogeneous has its utility for particular objectives of magnification or purification of factor pattern. But, in view of the dangers of these procedures, there is a great deal to be said, especially in the initial explorations of a field for working with populations which are normally heterogeneous and *as normal a sample of the general population as possible.* After the exploratory stages special clarifying studies can then well be made with such special manipulations of population. For example, in regard to research on abilities, after the exploratory stages in which the relation of abilities to other factors than general ability have been surveyed, we may naturally turn to the precise definition of the loading pattern, respectively, for males

and females, and for each of several years of age. And when all this knowledge is fitted together, the only important aspect of homogeneity-vs.-heterogeneity that remains to concern the psychologist is that he should remember to use the specification equation loadings that are appropriate to the degree of homogeneity in the group with which he is working!

Questions and Exercises

1. What features of design in factor analytic investigations are open to manipulation? Make a brief but comprehensive list.
2. In what ways are tests and persons not symmetrical (with respect to occasions) in factor analytic design? What are the implications for the value of Q-technique as an alternative to R-technique?
3. Summarize the ways in which choice of variables can influence the outcome of a factor analysis, and indicate what sort of choices are normally desirable.
4. Illustrate ways in which the actual units of population have differed among researches known to you. Discuss the manner in which the factors obtained in a research are to be interpreted in relation to the kind of population unit used.
5. What differences are likely to arise in correlation coefficients and factors from using as entries in the correlations the means of groups of size n instead of individual persons (as when we correlate population means for a collection of small communities instead of individuals measured on the same variables)?
6. Summarize the chief conclusions about the effect of alterations in the sigma and mean of a population, with respect to some or all tests, upon the factor resolution obtained.
7. Indicate the two chief reasons for the failure of a research to discover a particular factor (or to demonstrate it at a proper level of significance).
8. Discuss the pros and cons for artificially creating greater homogeneity or heterogeneity in a population instead of working with a normal sample. Indicate situations where each has real value.

CHAPTER **20**

Structuring Variables by Combinations of Factor Analysis with Controlled Experiment

In Chapter 1 we described controlled, crucial experiment, dealing with the relation of a dependent to an independent variable, as the extreme of one dimension of scientific method, and classical, non-interfering factor analysis (along with some other statistical methods) as the opposite extreme. Later it was indicated that many research designs used intermediates, combining physical control with complex statistical control.

SYNTHESIS OF FACTORIZATION AND EXPERIMENT

The discussions of the two preceding chapters have shown that even the classical factor analytic procedure *in situ* permits the manipulation of so many features of the design that already the beginnings of *experimental* control may be said to be creeping in. It is our purpose now to consider possibilities of combining factor analysis with higher degrees of experimental control and interference. Whereas the manipulations so far discussed involve only minor departures from passive statistical investigation and are likely to have been widely encountered by the reader in various discussions in classical factor analysis, the proposals now to be made are for a more radical combination of experiment with factorization. Neither classical experimenters on the one hand nor classical factorists on the other have yet shown much inclination to explore this hybridization, and in the absence of such exploration with actual experimental data, the reader must regard the propositions of this chapter as being of a more tentative nature than the routinely accepted and employed concepts of the main part of this book.

The argument that a more complete synthesis of the methods of

controlled experiment with the methods of the more powerful statistical tools can yield for many situations a far more effective methodology than either alone can offer necessarily implies that the preoccupation of the laboratory experimenter with traditional forms of laboratory manipulation and statistical design and expression must be viewed as an obstacle to research advance, even in what has been regarded as the strictly experimental field.

Let us first review the outcome of the general discussion of factor analysis in relation to scientific method undertaken in Chapter 1. The following main conclusions were tentatively reached:

1. Factor analysis has its most obvious role where controlled experiment is difficult or impossible, and the variables have to be examined in their natural situation.

2. Factor analysis is indispensable and without substitute in those early stages of a science where the natural functional wholes remain to be discovered in the chaos of multitudinous variables. For the factors representing these wholes are the dependent and independent variables which it is worth while to take under closer scrutiny either in subsequent controlled experiment or in more intensive statistical analysis.

The entities which the investigator with a broad approach most frequently wishes to relate as dependent and independent variables are themselves generally abstractions from a considerable set of operational variables. For those workers in biology and the social sciences who are not acquainted with factor analysis this has proved a baffling problem, which they have generally sought ineffectually to solve by taking a single "symptom" of the abstract concept in question, attempting to claim that if this behaves as predicted the hypothesis about the concept is correct. Nothing could be more misleading. The variance in any single, operationally defined symptom (dependent variable) is usually determined by many influences. The part due to the concept in question can only be determined by typing the latter down as a factor, by the other variables through which it is expressed. Thus, writing of some social consequences of the Oedipus complex, Winch exclaims (**154**): "Because they consist of high order abstractions, the major concepts of Freudian theory lack observable referents." But in his later work he recognizes that the problem is one of collecting and weighting the referents, not one of lack of referents, and the former is achievable by factor analysis. It has well been said that

psychoanalysis can be made scientific, not by experiment, but by becoming a branch of factor analysis. Generally speaking, it is a poor hypothesis—of the cheap variety formally imported because it is nowadays socially respectable to be clothed in a hypothesis—which can be tried out by appeal to one variable. A rich, well-thought-out concept, founded on patient observation, will generally be rooted in several variables and permit inferences as to combinations of relationships among them. Factor analysis is ideally adapted to testing theories extending to simultaneous relationships (patterns) among several variables.

The pursuit of intensive studies on single pairs of variables, even if guided by good hunches as to their conceptual reference, can seldom confirm a theory. At least in the social sciences the history of failure justifies designating this approach as "muddling" rather than "muddling through." The latter has sometimes been derogatorily attached to factor analysis in its groping for wholes. But an incisively designed analysis is far less in the realm of blind trial and error than is the practice of getting precise relations between two variables which are each complex in their factor constitutions and probably not very significant from the standpoint of the factor one is really interested in (as the loadings of later factorization sometimes show). But the real failure of the classical controlled experimental approach in these circumstances is not the lack of significance in the particular pairs correlated: it is the absence of all the other pairs of correlations which are indispensable to giving meaning to the first relation.

CRITIQUE OF METHODS OF STRUCTURING NEW FIELDS

The utilization of factor analysis in this exploratory stage of investigation is better labeled *the initial structuring of a field* by factor analysis. It is to be contrasted first with the above described attempts to obtain order by the choice of pairs of variables at random (or upon some blind hypothesis) leading to the working out of usually inappropriately meticulous mathematical functions to express one in terms of the other. This latter imitation of the physical sciences has also been attacked by von Neumann and Morgenstern (**98**) in connection with their advocacy of the type of formulation developed in the theory of games, a development quite different from but highly complementary to the factor analytic method. They write, "It is unlikely that the mere repetition of the tricks which served so well in physics

will do for social phenomena too," which ". . . should be remembered in connection with the current overemphasis on the use of calculus, differential equations, etc., as the main tools of mathematical economics."

But there are other attempts to handle this structuring problem without benefit of factor analysis which deserve as much criticism as the complex function equation on the grounds of inappropriateness. For example, in many studies in the social and biological sciences attempts have been made to handle the complexity of influences by use of multiple and partial correlation, as described briefly in Chapter 1. By the time as many partial correlations have been worked out as are necessary to eliminate this or that influence and, further, what is common to this or that influence, the result amounts to a patchworked factor analysis, but one in which it is possible to get lost easily in a maze of uncertainties. As to multiple correlations, their dependence on the sample and the obscurity as to what factors are actually accounting for the degree of prediction obtained have already been mentioned. Only in certain applied problems, where there is no interest in going outside the immediate system into the field of general scientific concepts, are such approaches reasonable.

Equally unsatisfactory as a substitute for factor analysis for truly scientific purposes is the use of the simple and the generalized discriminant function. This device shows how to combine, by appropriate weightings, a set of variables so as to get a maximum clarity of discrimination (reduction of overlap) between the members of two previously defined classes. It has been most used in determining, for example, what test weightings will best distinguish two or three occupational groups, the members of which have all taken the same tests. Little new knowledge about structure is gained, for the experimenter has to know beforehand what his groups are: the experiment does not discover the "clusters" as in factor analysis. Its use in the applied field of vocational guidance is limited by the facts (1) that one does not learn what factors—psychological source traits—are operative in the occupations and cannot therefore apply psychological laws regarding their change with age, training, etc.; and (2) one discovers nothing about the role of the factors (their regression coefficients) in *success* in the given occupations, but only the average endowment of those who do not move out of the occupation compared with those who do not move out of some other occupation.

These necessary observations do not detract in the least from the statistical brilliance of Fisher's development of the discriminant function and of its generalization by Rulon, Tiedeman, and others. It has indispensable uses in special—usually applied and emergency—situations, especially when the discriminations are made in terms of factors instead of mere variables, but plays a much smaller role in the development of scientific understanding of the nature of basic influences in biological and social data.

Some social sciences, notably economics, have attempted to structure the field by setting up a number of simultaneous equations. To make sure that the terms are really independent they are sometimes constituted from partial correlation coefficients from which the other influences have been removed, but this latter is uncommon. The formulation when the terms are not independent is inefficient because several terms may be expressions of one and the same underlying factor. In any case, this sort of equation usually presupposes that we know the direction of causal dependence between the term on the left and the terms on the right of the equation—indeed that the former is a sequel to the latter. One of the unfortunate results of uncritical imitation of the physical sciences is this assumption by the social scientist that he *knows*[1] the direction of causation in any correlation and that he is entitled to use the terminology of dependent and independent variables when in fact this conceptualization does not strictly apply.

METHODS OF EXPLORING CAUSATION AND INTERACTION

After factor analysis we are generally in a better position, with a given structuring of variables into factors, to ask by a further investigation which of these are dependent and independent variables. But initially, especially with complex social variables, we are not generally dealing with a true experiment in which we actually control one variable or have positive information that one is more basic in influence than another. The social scientist has tended to assume that the variable he thinks of first *is* the independent variable; but even when he is substantially right, it is very rare for some reverse causal effect to be absolutely impossible. Related to this is the mode

[1] For example, a renowned sociologist has recently reported before a Senate-House economic subcommittee certain discovered correlations between low income, delinquency, low sexual moral standards, etc. and has blandly argued that the correlation demonstrates poverty to be a cause of the psychological deficiencies observed.

of verbalizing which assumes some variables are endogenous or within a system, whereas others are exogenous. For example, in business cycle theory it may be assumed that inventions are exogenous; they affect the cycle but are not affected by it. This may be approximately true, but in good and precise methodological approaches it would be better to take a theoretical design which does not assume this.

In general not only do we lack information about a specific direction of causation, but we also have given to us a general directive from the whole of social and biological research to the effect that *most interaction will be circular*. The form of such interaction most commonly discussed today (**140**) is what has been called the feedback or servo mechanism. Change in a certain influence A causes change in B and the change in B instantly reacts back, usually negatively in a nonexplosive system, upon A.

Factor analysis is superior to most methods of exploration in comprehensively revealing such nexuses of interaction. It makes no assumptions about the direction of causal action, or about what is endogenous or exogenous to a system. If certain variables are in fact independent and outside the system, this will be shown by zero loadings in the factors that comprise the system. If nature does not know about the experimenter's favorite hypothesis which assigns pivotal importance to a set of supposed independent variables in a regression equation, the factor analysis will quickly show the fallacy of the supposed regression equation. If the experimenter has set up a criterion which he believes is influenced by such and such factors, the inclusion of the criterion in the factor analysis will quickly show whether in fact these factors need to be included in the regression equation for that criterion, and so on.

As to causal sequences, it is probable—though as yet it rests only upon a priori argument—that factor analysis can directly throw light on the sequence, at least when it is only in one direction. It would seem that in general the variables highly loaded in a factor are likely to be the causes of those which are less loaded, or, at least that the most highly loaded measure—the factor itself—is causal to the variables that are loaded in it. That is to say (when by independent we understand controlled), the independent variable is likely to be the factor, and the dependent variables are likely to be the variables. The argument is that the correlation of the factor with the variables is less than unity because it is attenuated by intervening events. For example,

temperature might be a factor in determining rainfall and many other things, but in each the operation of the temperature is modified by other influences which may intervene. The assumption (when we have nothing but the factor analytic evidence to go on) that the influence of the factor is attenuated in its affect on all the variables, is more simple and satisfactory than that all the variables are simultaneously attenuated in their influence on the factor. For example, when a general ability factor is found to load arithmetical performance, spelling ability, and good social habits, it is readily seen that these can be conceived as products of the operation of ability, the relation being attenuated by chance error and by the intervention of other measures which are themselves factors such as time and opportunity. Wider experience is necessary before we can generalize more confidently on this matter—before, for example, we can extend the argument to say that second order factors, organizers among first order factors, are also generally their causes.

The above argument applies to reasoning from a simple R- (or Q- or P-) technique result in which no specified time sequence observations are included in the data. Essentially all the measures in these are taken at the same time—even in P-technique. In the last resort of logic, however, the idea of causality boils down to nothing more than an invariable sequence in time, and *direct* evidence or confirmation of causal direction must therefore incorporate time sequence in the basic data. Factor analysis can utilize time data of this kind even when it cannot control variables, for it can repeat observations after time lapses. This is done in factorizations of increments and also in what may be called time-analyzed P-technique. The former has already been described; the latter means the employment of P-technique using staggered or lead and lag correlations. In staggered correlations the various scores on variable *A* are not paired with the corresponding scores on variable *B*, but (if the P-technique happens to be based on successive days for its score series) with the score taken, say, four days earlier on *B*. Let us suppose we are correlating the incidence of influenza with population movements and other conditions, and that the incubation period for this disorder is about five days. Then the highest correlation between the two series will exist, if this hypothesis happens to be correct, when the conditions series is placed five days earlier in phase than the influenza symptoms series, and will decline with a shift of phase in either direction from this position. [illegible]

tions could thus be carried out with correlations based on different amounts of lead and lag to see which give the clearest factor structure and the highest loadings. (Factor matching will of course have to be undertaken.) This will probably prove to be one of the most powerful means of exploring causal relations when using factor analysis aided by data employing a time signature.

QUANTIFICATION OF DISCOVERED RELATIONS

3. Let us now turn to developing more fully the indication in Chapter 1 that factor analysis also has a role at the more advanced stages of investigation where the initial structuring of the influences at work in the field, as in 2 above, may have been sufficiently accomplished, and where one is beginning to seek for precise quantitative statements of the relations among these influences. As pointed out in Chapter 1, the most common alternative statistical design, namely analysis of variance, fails to do more than indicate whether there is or is not some significance in the relationship between a supposed dependent variable on the one hand and several independent variables on the other. Factor analysis can be made to improve upon this not only in telling us from previous experiments or even the existing experiment what *independent* factors operate in our independent variables, but also in giving us a quantitative statement, in the form of a correlation or loading, of the *degree* of relationship between the condition or stimulus factors and the dependent variable. A correlation coefficient or regression coefficient, needless to emphasize, is not the last word in quantitative relationships; but it permits more precise answers to hypotheses and the testing of more developed hypotheses than does analysis of variance. Beyond the experiment to obtain a quantitative value for the regression coefficient lies indeed the further experiment or data collection required to plot a precise curve for the relationship expressed relatively coarsely in the regression coefficient, and perhaps to achieve the final aim of finding an exact mathematical equation to express the curve. As suggested earlier, an attempt at this final precision in complex fields like those of the social sciences is abortive or relatively meaningless until factor analysis has structured the influences, as described in 2 above, and given a first quantitative statement of the influences, as now being indicated in our continuation of 3.

This quantification in terms of regression coefficients, as done in

the typical specification equation, can be gained from any factor analysis. The above digression into an expansion of that comparison of factor analysis with other or older methods made in Chapter 1 was made in order to show that such quantification can be reached in realms sometimes not thought of as accessible to factor analysis. In fact, it is our purpose to show that the experimental and the factor analytic methods can be combined in syntheses of richer yield than is possible from either method in its simplest traditional area and manner of application.

WEAKNESS OF OPERATIONALISM WITHOUT FACTORIZATION

It is probably more instructive first to discuss, explore, and illustrate these combinations by a few particular instances and later to systematize the possibilities in briefer general statements. In the first place, there are many controlled experiments where the experimenter claims to be investigating the relation of one concept to another, e.g., anxiety and rate of learning, and where, though he has an operational definition of his concepts, there is no guarantee that they are as correct as they are precise.[2] For example, the naïve experimenter might plot the performance of a rat in a maze against the independent variable of strength of fear drive as measured in terms of voltage of electric shock administered. No criticism is being made of this precision, but if strength of escape drive is *not* best measured by voltage of electric shock, and if rate of learning is *not* best measured by the maze scores taken, then the conceptualization is wrong and the experimental findings have no reliable wider application. The design here advocated would be to measure several

[2] The operational definition is a precise enough definition—of the operation. But except to make the experiment reproducible—and an exact description of the experiment will do this—the definition must deal with something beyond the operation. It must have reference to a concept *behind* (abstracted from) the operation if it is to have wider reference and usefulness, e.g., in making possible relations between the concept and other concepts. Usually *one* operation is as inadequate to tie down a concept as is one variable to define a factor. Unless the variable is a *pure* factor measurement, it has other factors in it than the one referred to, and the factor has other variables to express it than the variable in question. For example, the strength of the thirst drive may show itself by frequency of passing a punishment obstruction to get water, but thirst is not the only factor in this frequency (**30**, page 198), and other variables, e.g., length of waiting at a barrier, are also highly loaded with the thirst factor. Consequently, to define strength of thirst operationally by one of these operations, instead of by a combination of all of them, is arbitrary and inadequate.

variables deemed expressive of fear and several of learning and to factorize the complete data obtained in the usual learning experiment. Thus the first research step would be to find what factors are at work and what variables need to be weighted to estimate them correctly in order that the curve expressing their relationship may be reliably plotted. A more specific aspect of this general argument for proper representation of both dependent and independent variables occurs in some recent criticism of analysis of variance designs. It is rightly pointed out that the experimeter is always concerned to look for significant differences in the dependent variable without being careful to get significant differences first in the independent "effects." How representative are the differences in the latter of the differences normally occurring in nature? In factor analysis the representation of the "independent" factor by several variables, and the demonstration that they have significant common variance, by the very fact that a factor appears, is a guarantee of significant variation in it.

But the main issue is that the variables should truly represent the influence hypothesized. Thus in an investigation of the effects of fatigue, the experimenter, imitating the physical sciences, might put his subjects through exhausting work performances and then measure the decay through fatigue of various dependent variables instead. The factorist would advocate analyzing the group of *supposed* fatigue symptoms, which might well lead to two or three distinct varieties of fatigue being estimated as separate factors and to the separation of two or three corresponding curves of functional relationship previously confused in one. The operational approach makes complete sense only when factor analysis is added to determine the generalizable referents of the exact operations.

EXAMPLES OF COMBINATION OF METHODS

A combination of factorization and experiment quite different from the above is that in which an experimental variation of conditions is introduced which may affect *all* of the variables. For example, the intercorrelations of a set of price variations may be factorized first in conditions of free competition and then under some condition of government subsidization. The difference in the nature and number of factors in the two conditions could present a crucial test of some economic theories. In such a design, the question will occur to some as to whether the alteration of the total measurement conditions from

one factorization to another forbids one to consider the variables as the same tests on the different occasions. For example, one might give two dozen ability tests under ordinary conditions and again under speeded or highly motivated conditions. This undoubtedly alters the nature of the test variables by any accepted definition of test. But the same is true of any experimental psychological investigation of a relation between a dependent and an independent variable; the mental operation involved is likely to alter at different parts of the curve as the circumstances in which the test is given alter systematically (constituting the controlled variable). Indeed we encounter at this point a question of formal symbolization of the stimulus which has never adequately been treated in personality and learning theory and which must be cleared up in order properly to formulate the designs shortly to be described. Briefly, the idea needing recognition is that every total stimulus situation is a situation within a situation e.g., a test, duly defined, within a set of more general conditions which also need defining. That which is extra to the characteristics of the organism, i.e., what is roughly called the environment, is not to be defined by a single measurement, even when the attention of the organism is on a single aspect of the environment. It has a whole series of dimensions.

The general psychological theory implied, in whatever present discussions on factor analysis hinge on psychology, has been that expressed by the present writer elsewhere (**22**) (**30**) and which differs from reflexological, stimulus-response formulas in writing

$$R=f(OS)$$

i.e., the response is a function of both organism conditions and situational conditions. This integrates historically the McDougallian dynamic psychology with factor analysis, in that the latter has necessarily been closely concerned with the organism variation, whereas the experiment on the design of the physical sciences, using for example, analysis of variance, has tended to neglect this as error. The factor analytic specification equation already breaks down the description of the stimulus situation into a whole set of situational indexes as we have seen. But the present need to break down the stimulus goes beyond that. The situational indexes are various aspects of an intact situation. The present proposal, on the other hand, is not to leave the stimulus situation as static and intact, but to vary

the different conditions independently and explore the effects of such partial variations upon the factor analytic picture. One way of doing this directly (but which will not be followed up here) is by T-technique (see page 110), which gives a veritable factor analysis of the dimensions of the stimulus situation. The beginning of such a splitting of the stimulus—at least to the extent of regarding quantitative effects as due to differential action—is already evident in the parallel field of learning in the formulations of W. K. Estes. For the moment we shall leave the problem at this level of emergence and explore further general designs.

Yet another hybrid of factor analysis and experimental control which has been suggested involves P-technique. In this, the variation of conditions with the occasions axis in P-technique is deliberately controlled instead of being left to chance. For example, in an experiment to determine the pattern of drives (ergs, **30**) in man, the individual might be tested from day to day on the strengths of a variety of interests and attitudes hypothesized to involve hunger, the sex drive, fear, etc. The stimulus conditions for any one of these drives could be deliberately varied from day to day, as also could the inherent reactivity of the individual, by experimentally inducing physiological change. Only the uncontrolled, unmodified factor analytic design has yet been applied here (**23, 32, 34a**).

SYSTEMATIZATION OF POSSIBLE COMBINATIONS

To systematize the possibilities of new designs as now briefly surveyed above, we need to ask about the basic nature of the observations in factor analysis. An experiment is an investigatory situation in which the experimenter does some precisely defined thing to an object of investigation and observes what *it* does in turn. For example, he may heat red lead and record what happens or drop a stone under defined conditions and time its fall or raise the temperature of a new chemical substance to observe its melting point. Unquestionably in the great majority of true experiments, whether in physical or social science, the essential operations or elements are two—presentation of defined conditions and recording of observed events. In psychology, for example, this is presentation of a defined test to an identified individual and measurement of his response. Some experiments may seem at first not to fit into this paradigm, as when a chemist says, "I am producing metallic uranium to see

what color it is." The discovery type of experiment may properly be regarded, however, as an extreme case of the above type, wherein the experimenter's conditions are not very salient or explicit. (The above question should strictly be, ". . . to see what color uranium as produced by this defined process turns out to be when viewed by white light.") Incidentally, it is a matter of indifference to the primary definition of an experiment to state whether a prior, explicit, and detailed hypothesis is involved or whether the hypotheses is simply that something will happen.

Granted that an experiment when more closely examined falls into the essential form which we have tacitly assumed all along, namely, the seeking of a relation between an independent, controlled variable—the imposed conditions—and a dependent variable—the response or observed change; how can this mesh with the system of observations in factor analysis? The basic set of observations in factor analysis, as schematized by the covariation chart (page 109) shows that each measure is tagged by three referents (actually five, as shown in Chapter 8, but only three are commonly variable). They are a person (or organism), a test or response measurement, and an occasion (or condition). Incidentally the actual score matrix is a two-way, not a three-way score matrix because in any correlation series one of the three is held fixed and therefore not mentioned.

Now in this scheme the conditions of the stimulus situation, though defined both by the test and the occasion (the former defining the narrow stimulus, this latter the broad situation) are held constant—except in so far as accidental, nonsystematic variation enters into the occasions axis. But if now we set out deliberately to introduce variations of stimulus conditions we might regard the new axis of variation as a rib taken either from the body of the occasions axis or from the test. It seems more correct to consider it as a function of occasions, for in almost all controlled experiment there will be some aspect of the stimulus situation intimately associated with the response measurement—which indeed may be considered a defining feature of the "test" or response measurement—and which does not alter as the main conditions are systematically altered. Consequently in this wider generalization the axis we have called tests will have its emphasis shifted a little to mean *nature of response measured*, while the occasions axis will now be frankly a *systematic* change of overall conditions with occasions. And since the definition of a stimulus

situation, as we have seen above, really requires attention to *several* distinct parameters or aspects of the situation, this new conditions axis will actually split into several axes, as shown in Diagram 32, corresponding to the conditions of the stimulus situation which may be independently varied in controlled experiment.

These axes are spatially to be considered as fourth, fifth and higher dimensions to the covariation chart, though this cannot be indicated more than symbolically in the drawing of Diagram 32.

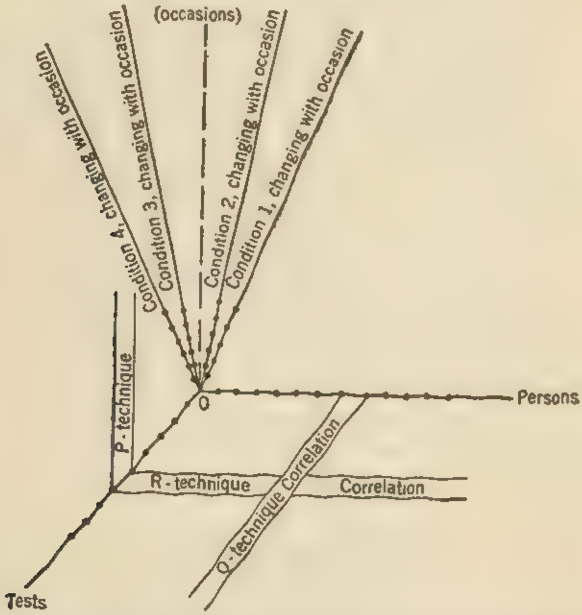


DIAGRAM 32. New Dimensions Added to the Covariation Chart when Controlled Experiment Is Introduced.

Now a correlation is always represented in the covariation chart by a pair of parallel lines utilizing two dimensions while the third is held constant. Consequently these new dimensions mean a host of new possibilities of correlation, e.g., one test could be given under two states of condition 1 to a series of persons; one test could be given under two states of condition 1 to one person while condition 2 is varied over a whole series of positions throughout its range and so on.

So far this introduces controlled independent variables into factor analysis but does not change the classical factor analytic design, for it still takes differences in only two things at a time and forms the usual correlation series therefrom. Thus all the various possibilities of covariation can be explored only in a succession of factor analyses, holding everything but two dimensions constant while each is proceeding.

Out of a greater number of theoretical possibilities of combining controlled and uncontrolled variation we shall now proceed to describe in some detail five major designs which seem to have practical importance. The first are of a more general nature, and it is not until we get to (4) and (5) that the special developments just discussed in regard to Diagram 32 become highly relevant.

FACTOR ANALYSIS OF INDEPENDENT AND DEPENDENT VARIABLES IN A COMMON MATRIX: CRITERION ANALYSIS

In this design, already illustrated above, the changing *conditions* and the changing *responses* with respect to several different variables are correlated (for one person) and factorized. The design is analogous to P-technique, indeed it is controlled P-technique. It is also the classical experimental design, but the two tests correlated are now, in half the cases, the test responses on the one hand and the test conditions on the other. If several test conditions are used they will need to be randomized with respect to one another, in relation to occasions. Its advantage is that no assumption is made, as in the classical controlled experiment (a) as to which variable is really dependent or independent and (b) as to the unitariness of either the imposed influence or the resulting dimension of response. For example, in the above suggested experiment on fatigue several distinct fatigue-producing conditions could be simultaneously tried and several distinct supposed measures of fatigue effects could be employed. Distinct fatigues and distinct groups of fatigue producers would thus be isolated for further experimental work on the precise mathematical form of the relationships. These relations would thus be the first indicated by rough correlations and later by exact graphing.

The series for the correlations in this design can be visualized in Diagram 32, as composed by one of the P-technique verticals on the one side and one of the conditions verticals on the other—the condi-

tions being considered independent of any test—erected on test number 0.

The design of including the dependent and controlled variables of a classical "experimental" type of investigation in a factor analytic instead of a factorial (or "difference of means") design—and which we have called below, in its widest sense, *condition-response factorization (or analysis)*—includes as a special case what has been called *criterion analysis.* In recent years several psychologists concerted to factorize a set of tests *and* find the regression of the factors upon an external criterion have found it convenient to include the criterion in the test matrix from the beginning. In this way the criterion is "analyzed," in the sense that we know—at least for that sample—its factor composition, i.e., the specification equation by which "success" therein can be predicted from factors and ultimately from tests. Incidentally, it will be doubly evident at this point why we have suggested that "criterion rotation" is a better term than "criterion analysis" for the rotation procedure of Eysenck described in Chapter 14. Criterion rotation artificially makes the criterion correspond exactly to a factor and therefore does anything but analyze it!

The convenience and statistical neatness of criterion analysis has so far not been shown to be offset by any systematic vice in the design. There are, perhaps, two points needing to be watched: that the regression of the factors on the criterion should be checked on another sample, not using factor analysis necessarily; and that the possibility of curvilinear relations to the criterion be investigated, for whereas tests nearly always relate linearly to factors, factors sometimes have an "optimum" value in relation to the criterion. Where the latter holds, a coefficient of pattern similarity (26) is a better measure of success on the criterion than a regression equation.

Either in the special case of criterion analysis or in the general case of condition-response analysis, where any two "classes" of variables can be considered, the outstanding advantage is in being able simultaneously to deal with several conditions and several responses. Thus in criterion analysis we can throw in several criteria, each of practical interest for the vocational or clinical selection concerned, without the incorrect assumption commonly made that they are all equally measures of "success." Indeed the factorization may incidentally show that there exist two or three independent criterion factors.

The above considerations, arising out of the third general statement

on factor analytic design in scientific method—that it is applicable to the more advanced research purposes of determining quantitative relations of variables beyond those of more "statistical significance"—remain to be formulated in a fourth statement as follows:

4. Factor analysis can be used with experimental control of conditions (independent variables), to produce more extended knowledge of structure and meaning than is obtainable from classical dependent-independent variables experimentation.

Two main kinds of hybrid of controlled experiment and factorization will be considered below: that in which conditions are subject to controlled change between one factorization and another (next section) and that in which different blocks of variables in the one factorization correspond to different controlled conditions.

FACTOR ANALYSIS OF DIFFERENCES BETWEEN TWO FIXED CONDITIONS

The possibilities to be considered systematize a design already given in illustration and a design already used in an actual psychological problem (**22, 143**). Since there is here no question of simultaneously varying two or more conditions, the possibilities can be seen from a simplified version of Diagram 32, as given in Diagram 33, in which there is only one condition axis.

First, we may factorize (using R- or Q-technique) under one fixed condition and infer laws from the relation of the results to those of a second factorization under new conditions. An example from ability factorization has been given, but others would be personality test structure with and without alcohol, group syntality with and without authoritarian leadership, or dynamic investment patterns of personality before and after psychotherapy.

Second, we may make a *single* factorization in which the scores are *differences* between the responses under two different conditions. Under the title of factorization of increments this has already been sufficiently illustrated (**143**) and discussed (**22**), but there are a great many situations in which it remains to be applied. For example, one might measure change in performance in a number of learned activities after the interposition of an influence designed to produce retroactive inhibition. The result would indicate whether one or more influences operate in transmitting the retroactive effect. The increment series for correlation are shown in the base of Diagram 33.

In both designs, a modification is possible in which the controlled influence could be one brought to bear upon all variables in the battery or could take the form of several influences each brought to bear on one of a block of overlapping or nonoverlapping subgroups of variables. The factorization would then test the hypothesis that these are indeed independent and distinct influences.

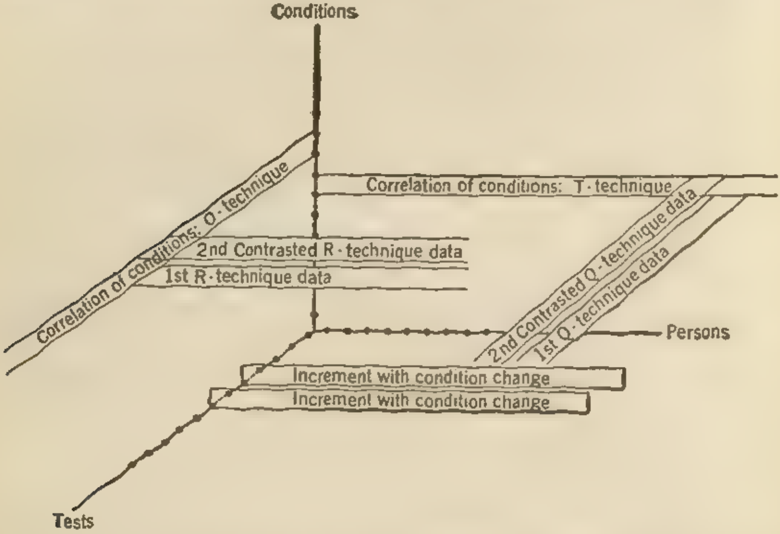


DIAGRAM 33. Correlation Series in Covariation Chart for Factorization of Experimentally Produced Changes.

FACTOR ANALYSIS OF CONDITIONS

What has been briefly referred to as correlation of occasions (22) with particular reference to history and personality history becomes, with controlled conditions substituted for occasions, a tool of wider utility. The paired lines for such correlation series, using T- and O-techniques, are shown in Diagram 33 and could be shown for P-technique in a fourth dimension of Diagram 32. Typically (T-technique) one would measure a certain test response for a whole series of persons under condition a, again under condition b, again under c, and so on. The correlation matrix would be one ready for factorization of the stimulus conditions a, b, c, etc. There is an analogy to analysis of variance in one respect: that only one dependent variable (the response) is measured and several different independent vari-

ables are applied. If in the equivalent analysis of variance experiment no significant interrow differences were found, that would mean no test reliability or correlation. The factorization should reveal how many distinct conditions are actually being varied in all these apparently diverse stimulus conditions, and should lay a foundation for better experimental work economically concentrating on the truly independent dimensions of the stimulus.

CONDITION-ORGANISM (OR CONDITION-RESPONSE) FACTORIZATION WITH SEGREGATED CONDITIONS

It has been pointed out that the designs now being considered, particularly numbers 4 above and 5 below transcend the preoccupation of classical factor analysis with variation inherent in individual differences and include variation in the controlled stimulus conditions. The particular virtues of such a design include (a) that we may obtain a sense of proportion, perceiving how much of the variance in response is due to environmental and how much to organismic variance, and what nature and number of factors need to be taken into account in the whole problem; and (b) that we may combine the excellences of controlled experiment with the analytical power of factor analysis. The correct use of (a) will require that we use a magnitude of "controlled" variance similar to that occurring in nature.

We shall first approach the possibilities through design 4, systematizing what is discussed in the above section, and which is a partial, segregated form of what becomes complete in design 5 below. Essentially a certain condition is applied to the population while they are performing each of three or four tests. If this is a significant influence in the performances covered by these tests, it should produce correlation among them additional to any that might already exist and should yield a single factor—the influence itself—running through these tests. Whether or not we throw into the factorization a variable corresponding also to the stimulus condition as it was imposed (scoring each individual according to the strength with which the influence was brought to bear on him) is a matter of further choice in design, but we shall assume here that the greater factor definition obtained by including it favors its inclusion as what we shall call the *condition variable.*

To take a current psychological problem as an example (33), let us suppose we have a hypothesis that rigidity or resistance to learning

(in a wide variety of learning situations) is variously due to four factors as follows: F_1, or intelligence (negatively); F_2, or strength of motivation (negatively): F_3, degree of fear or apprehension (positively); and F_4, native temperamental disposition rigidity or *p* factor (perseveration) (positively). By hypothesis and existing indications factors 1 and 4 are inherent in the individual and can be left to vary naturally in the experiment. To reveal these factors, it is necessary only to introduce among the nucleus of variables, say V_1, V_2, V_3, and V_4, which are concerned with various ordinary forms of learning; a couple of variables, V_5 and V_6, which are specifically markers for intelligence and largely concerned with relation-education learning. Also we shall add a couple of markers, V_7 and V_8, specifically known to be highly loaded in the classical disposition rigidity or perseveration factor, e.g., a motor test writing letters with reverse strokes and a hidden pictures perceptual test.

But, in contrast to F_1 and F_4, the supposed factors F_2 and F_3 lie in the *conditions*,[3] and the way to demonstrate them is to take one subgroup of the existing tests—say, the odd numbers V_1, V_3, V_5, and V_7—and arrange that people do them with varying degrees of motivation, Roberts always having the strongest incentive offered him, Jones the weakest, and the others of the population in a fixed order between.[4] Similarly the second applied stimulus condition (fear) could be arranged, with respect to tests V_2, V_4, V_6, and V_8, in a common rank order of strength with respect to individuals. This order would be best arranged to be random (orthogonal) with respect to the order of the first stimulus order, maintained through V_1, V_3, V_5, and V_7, though this is not essential, for we can deal with correlated factors. Again there would be a pure condition variable, V_{10}. The fact that a

[3] The recognition of diverse factors in the conditions of the stimulus situation as well as the splitting of the occasions axis into several conditions in Diagram 32 is, incidentally, in accord with the dawning recognition in learning theory work that the stimulus situation can no longer be expressed by a single term in the learning equation, but must be broken down into several stimulus elements. These distinct dimensions of the stimulus may turn out to be functionally related to the situational indexes in the specification equation.

[4] As a helpful refinement, we can vary the *range* of the condition as between each of the different variables in which it appears. For example, the *range* of the motivation could be greatest for test 7 and least for 1. We should then expect that if a factor appears in V_1, V_3, V_5, and V_7, it will be highest (if the condition is otherwise equally potent in affecting all performances) in V_7. It should presumably be higher still in the new condition variable, V_9, which, as stated, is simply *the order of the population on the motivation (stimulus) condition itself.*

single influence has been supposed operating through each of these subgroups does not prevent the loadings associated with the condition appearing eventually divided into two or more factors if two or more actually exist (though a further search will then be necessary with new hypothesis variables to discover how they differ).

The design for factor analysis of imposed block variation of conditions can be summarized for the present illustration in Table 39. Here the variables are numbered as stated above and C_1 and C_2 are the two stimulus conditions (respectively, motivation and previous training) which are broken down into 12 steps and randomized with respect to one another, i.e., treated as in a Greek square or in a higher order classification (confounding of influences) in analysis of variance. Indeed it will be seen that we are dealing with a hybrid of factor analysis and analysis of variance, in which the rules and calculations of arrangement of cells as worked out for the latter (see, e.g., McNemar (**95**), Chapter 14) need to be followed.

If the conditions and variables are of a kind which permit it, the two blocks can overlap in those variables to which both conditions can be simultaneously applied. This may not always be practicable and, if carried to complete overlap, it also creates some difficulties in identifying the factors which emerge. A complete overlap of influence blocks is in fact considered as a separate problem of design under design 5 below.

The factorization of the twelve variables listed for the illustration just given should yield four or more factors: one high in the markers for general intelligence and variously loaded in all the others; one high in the markers for native rigidity and otherwise variously distributed; and two in the odd and even variables, respectively, corresponding to the rigidity producing conditions of low motivation and fear. These should be highest, as indicated, in the variables specifically defining the degree of the condition and, other things being equal, in those response variables in which it has been expressed with highest range.

If this factor analysis of imposed variation of conditions is examined in terms of the covariation chart (extended by a condition axis), it will be seen that the correlated series run as in Diagram 34. Here the dimension of occasions is held constant since all tests are on one occasion. It will be seen at once that the form is radically different from any previous analysis of a factorization plan on the

covariation chart in that the series run diagonally, indicating a prearranged interaction between two dimensions—persons and conditions in this case. (Design 1—Independent and Dependent Variables in the Same Matrix—may involve this also with respect to occasions.)

If more than one condition is involved, as in this example, it is desirable to orthogonalize (randomize) the scores on the condition variables relative to one another, when confounding the "effects" in a way familiar in Greek square designs in analysis of variance where

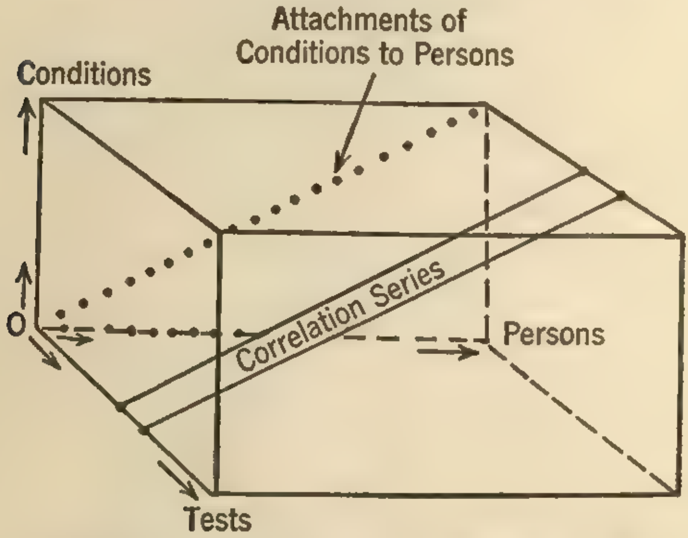


Diagram 34. Correlation Series in Covariation Chart for Factorization of Imposed Variation of Conditions.

the similar independent variables are so treated, as indicated above. The same rules apply to the number of cases required to get instances of all possible interactions for a given number of variables. If one is going to divide the range of each variable into twelve divisions to get the best use of a product moment correlation, one needs a population of 144 persons, or a multiple thereof. However, if several conditions are to be invoked, the population required can be kept small by reducing the number of divisions on each and ultimately by using tetrachoric coefficients.

The only defect of the design resides in the loss of interaction between condition and population due to an arbitrary assignment of

conditions to individuals. For example, if this correlation were extent of alcoholization imposed in relation to several performances, an interaction which might be called natural susceptibility to alcohol effects might not separate itself as a distinct factor from extent of alcoholization. But if some variables outside those covered by the alcohol condition happen also to be affected by susceptibility, then its splitting off as a separate factor will be aided. However, unless some factor chances to be ruled out by the experimental arrangements in this way or by omission of variables, the factorization will achieve its objective of surveying the number, nature, and magnitude of the stimulus condition and organismic factors operating in the area of behavior represented by the variables. Thus, the investigator is rescued almost forcibly from the abortive formulation of classical stimulus—response psychology on the one hand or the restrictions of classical factor analysis with all conditions held constant *in situ* on the other.

Incidentally, as mentioned in an opening sentence, whenever an R-technique design has been described, a corresponding Q-technique innovation is also implied. P-technique can similarly be inferred from designs 1 and 2, but in designs 3 and 4 it requires some explanation for the measures must now differ in respect to a *second* condition, i.e., a condition other than that which spontaneously varies through the series of occasions. The latter *may* be largely an internal characteristic of the organism, but it may also be an external condition of the same class as that which we experimentally apply.

CONDITION-ORGANISM FACTORIZATION WITH GENERALIZED CONDITIONS

The chief reasons for segregating each condition to a particular block of variables in design 4 are (a) that it may be physically impossible simultaneously to fix particular values of different conditions in one and the same experiment as required in the nonsegregated, and (b) that it may be more difficult in the nonsegregated to identify factors because each factor is now likely to be spread over all variables. However, if the former can be overcome, the latter can be met by invariably including the condition variable and by varying its standard deviation in different tests. Needless perhaps to say, the conditions must also be such that they are meaningful for all variables which they cover. For example, one stimulus condition might be a

TABLE 40.

Subjects	C_1	C_2	C_3	V_1	V_2	V_3	V_4	etc.
Roberts	4	1	1					
	4	2	1					
	4	3	1					
	4	4	1					
	4	1	2					
	4	2	2					
	4	3	2					
	4	4	2					
	4	1	3					
	4	2	3					
	4	3	3					
	4	4	3					
	4	1	4					
	4	2	4					
	4	3	4					
	4	4	4					
	3	1						
	3	2						
	3	3						
	etc.	etc.	etc.					

range from speeded to unspeeded test conditions, and here it would be pointless to include a standard reaction time test among the variables.

The advantage of design 5, if it can be used, is that it compactly permits more response and condition variables to be included in a factorization of a given size. Its disadvantages are the restrictions just mentioned upon its use—the restrictions upon the freedom of the experimenter in carrying out the experiment. He has to meet a lot of conditions by careful prearrangement, notably in orthogonalizing of the conditions, in keeping tab of the sigmas of conditions allowed to operate in different variables, and in insuring that the proper combinations are applied to each subject. With many variables this amounts to requiring an individual testing of each subject, for only one may fall in each cell of that double classification which, as indicated in design 4, we borrow from analysis of variance. (Incidentally, the intraclass correlations in the more complex analysis

of variance designs are a reaching out toward this design in factor analysis, though still remote.)

The score table for a three-condition, four-response variable design of this type is shown in Table 40 in which four grades of the condition score only are used and randomized among the three. Each variable will now have all three of these scores, e.g., Roberts' V_1 performance will have stimulus conditions $C_1.4$, $C_2.1$, and $C_3.1$ applied to it.

Our fifth and last generalization about factor analysis in research methodology may in conclusion be set out as follows:

5. Factor analytic designs are feasible in which pure measures (markers) for various applied conditions are correlated in with the variables expressing performance under varying controlled variances of these conditions. The design is a "hybrid" with analysis of variance in that the incidences of conditions upon individuals need to be confounded like effects in double classification analysis of variance. Its extra potency resides in yielding the degree of association, the structure among dependent and independent variables and the possibility of dealing with more than one dependent variable.

The above methods, being in their infancy, need further examination both as to their empirical effectiveness and their statistical conditions, but they offer such great promise as flexible and comprehensive designs capable of yielding quantitative answers and combining experimental control with sophisticated analysis of influences, that they justify description at this stage.

Questions and Exercises

1. Describe the role of factor analysis in the initial structuring of a field of scientific investigation. At what other stage is it particularly valuable?
2. Indicate the limitations to attempts to structure a field of variables by (a) exact work on the function expressing the relation of particular pairs of dependent and independent variables, (b) partial and multiple correlation, (c) breakdown into exogenous and endogenous variables without assumption of feedback effects.
3. What are the dangers of assumptions about directions of causal action before a field is completely investigated? Indicate three examples of sets of variables in which interaction is complicated by servo or feedback mechanisms.
4. What light can factor analysis throw on causal relations (a) as when it is instantaneous, static, without time data as in ordinary R-technique and (b) in P-technique with staggered correlations?

5. Indicate some limitations and even fallacies in the conceptual scheme of the classical pure experiment which can be remedied by factor analysis. Why is an operational definition of a concept not sufficient to define it?
6. List, with very brief descriptions, five hybrid research designs combining factor analysis with experimental manipulation and control.
7. Describe the condition-organism factorization design for analyzing organism attributes and imposed, randomized multiple conditions. What are its advantages and limitations? What would be the difference between an R- and a P-technique design along these lines? Compare the design and its objectives with that of analysis of variance.
8. Describe how the three-dimensional covariation chart of classical factor analysis becomes extended when systematically representing the possibilities of combining factor analysis with experimental control. Can you indicate from the extended covariation chart at least one theoretically workable design not mentioned in the five above?

CHAPTER 21

Strategy and Tactics of Economy in Computing

If social scientists are to be as realistic about the history of their own methods as they aspire to be about their subjects of investigation, they must admit that the failure to apply factor analysis as early and as widely as necessary has been due to irrational reactions unrelated to its real scientific usefulness. Older investigators have looked askance at having to acquire skills in what appears to be a complex brand of mathematics, and younger ones have shied away from the threat of so much sustained labor as the method seems to demand.

LABOR SAVING IS THE PRIMARY NEED

To individuals who so react one must point out that the complexities have not proved beyond the scope of a volume of this limited size nor beyond the capacities of any person capable of genuine scientific work. It may be necessary also to point out, as in the last chapter in earlier discussions of clusters, that the alternative approaches which seem to do the same thing in an easier way are not in fact capable of delivering the goods. Most partial and multiple correlation directed by expediency is nearly always labor lost as far as generalizable scientific advance is concerned. And searching for clusters instead of factors is to follow a will-o-the-wisp into a quagmire of unexpected and unprofitable toil. Conceptualization in terms of factors produces stable results of which these short cuts are not capable.

To the reaction that the actual computations are excessive we are

now fortunately in a position to reply that this once ominous observation is no longer true. The increase in machine aids and the exchange of tricks of the trade among the many craftsmen now engaged in factor analysis is rapidly leading to a repertory of laborsaving devices which enables any skillful performer to attack problems that would have appalled his predecessor of fifteen or twenty years ago. For example, most studies in the early days of multiple factor analysis handled about twenty variables and half a dozen factors, whereas now about sixty variables and a dozen factors constitute a more common and more easily handled assignment.

It is the purpose of the present chapter to concentrate upon the facilitation of computing by all possible aids, both by suggesting new devices and by summarizing those scattered through earlier chapters.

Some of the quicker devices call for higher levels of experience and skill in the computers, so that the designer of an experiment may still be well justified, in view of his available help, in planning to follow some of the more plodding methods. It is certainly preferable to be safe and sure in one's results than to take any short cuts involving possibilities of gross mistakes or approximations of unknown magnitude. Some of the aids now discussed are therefore understood to be contingent upon the level of skilled help and the types of calculating machine available.

There are essentially four stages at which timesaving methods can be introduced, namely, (1) in the design of the whole research (in which we may include the question of testing time or data gathering, extraneous to our present problems of computing), (2) in the computing of correlations, (3) in the extraction of factors and (4) in the process of rotation. Except for the first, which has reference to all, it is possible to consider these independently.

RESEARCH DESIGN

The problem of research design is, of course, one of overall research goal attainment in relation to time spent rather than of labor saving in the narrow sense. As to what we can manipulate, let us remind ourselves that except for the control of conditions mentioned in the special experimental design of the last chapter, the experimenter can control principally the kind and number of his variables and the

kind and number of his population as well as the relation of these features of his research to those of any other existing research.

Let us first see what economies are possible with regard to variables and then turn to population. With regard to variables the experimenter will aim definitely to exceed the minimum number required to define (page 334) the number of factors suspected to exist, and he will be guided in choice by some such concept of trait population as the personality sphere as well as by the need to introduce well-known markers (at least two and preferably three per factor) from previous researches, and the need, perhaps, to pack variables densely with respect to one or two factors sought with a high degree of definition. Naturally, one would avoid the waste of having two very similar variables involved which are unlikely to add to the meaning or fixing of the pattern, unless they are important for some practical purpose.

It is economical, in regard to further research orientation, to plan for each of the variables to be split in such a way that a definite reliability coefficient can be worked out, in order that loadings may be subsequently examined for significance, or corrected for attenuation, but the question of whether to work with long tests of high reliability or short ones of low reliability will depend on the research objectives. Where the subject's time is limited and it is desirable to explore a large area of performance, a first research may best use distinctly short tests, since the required factor structure will appear even though no (uncorrected) loading on any factor is really high.

It is generally advantageous to have not too great a variation of length and reliability among variables, for a correction is always only an estimate. It perhaps goes without saying that reliability and practically everything else is improved by taking some care—even in first explorations—to insure that variables are of about the right degree of difficulty and capable of giving a generous standard deviation. A J-shaped distribution does not invalidate the use of the product-moment coefficient but it does block the use of such computational short cuts as the rescaling method of correlation mentioned below.

Where research objectives demand that a rather large number of variables be used—and this is most likely to be a common occurrence in educational tests—and where the factor composition of a large

number of single-question items[1] is required, it is worth while to plan beforehand for a splitting of the correlation matrix into two or more parts, overlapping or clear according to the plan for later joining (see below).

COMPUTATION SAVING BY CHOICE OF VARIABLES AND POPULATION

The work of computing correlations and that of factor extraction are proportional (approximately) to the *square* of the number of variables involved. Consequently a considerable saving is achieved by breaking down a matrix of, say, 120 variables into two matrices of 60 each. Unfortunately some fraction of the saving is bound to be lost by the morticing necessary to put them together again; for if they are treated *simply* as two separate factorizations, part of the possible information is lost, notably in that one cannot tell which of the factors in one corresponds to each of the factors in the other.

There are four possibilities for establishing cross reference as

[1] Here the problem arises of whether first to item analyze small blocks of items, each block being then used as a variable in factorization, as Loevinger suggests (86), or whether to factorize items directly. The psychologist and the sociologist are not involved so frequently as the educational psychologist in measurement situations with pencil and paper tests involving yes-no or multiple-choice responses to a set of items individually equivalent save for test meaning. Nevertheless the problem is general enough to justify some further discussion here. Since a test of reasonable reliability usually requires a total of some hundred or more items, the overwhelming preponderance of research on such tests has followed the methods of item analysis rather than factor analysis. Yet the argument for factor analysis is as cogent here as anywhere else.

Let us glance briefly at what item analysis does. The items are correlated either with an external criterion (requiring only as many r's as items) or with an internal criterion constituted by a majority of the items or by all the items. The two last procedures may require as many r's as the factor analysis, but demand much less work later on the r matrix. None of these procedures purifies the test factorially. The first two intensify a particular conglomerate which, like variables weighted from a multiple r, gives the best prediction of the criterion at that time, but it is open to most of the objection facing a multiple r. The item-analyzed group of items remains of quite unknown psychological composition and meaning. The correlation of items with the pool yields an elect which is not stable; it is a method which sets us out on a very long chase. For when the items having low r's with the pool of all items are thrown out and the experiment is repeated, the remaining r's change their order of goodness. With repeated correlations and purgings, one factor in the conglomerate is bound at some point to gain a lead over the others in its mean loadings of all the items, and as soon as this occurs, it will forge ahead with increasing rapidity until the last few purges reach that stable position in which the items are relatively pure factor

follows: (1) One can simply allow the two groups of variables to overlap with respect to, say, twenty variables (i.e., have two matrices of 80 each) hoping that chance or previous hunches about the factors will permit the particular choice of 20 overlapping variables adequately to define all the factors. (2) One can factorize one matrix first, then estimate for each individual his possession of each of the factors by adding scores for the loaded variables and correlate these *factors* (perhaps, say, 10 of them in all in the case cited) with the variables of the second matrix. (3) One can correlate the variables in the second part of the battery with those in the first but not with one another, i.e., avoiding a complete second matrix. The first is factorized and the loadings of the variables of the second part in these factors are determined by means of Dwyer's extension method (44) or the shorter method explained below. This method is most used when the second part—the extension—is smaller than the first. (4) One can factorize one matrix first and then, instead of taking say 20 variables at random to carry over into the second as in device

---

measures (or have, at least, no second factor in common). In the special case where a large number of items have a very similar but complex factor composition the process will stop at that complex position, short of factor-pure items. But only a small fraction of the original pool is likely to remain in this factorially pure or factorially highly similar group, and the process of getting rid of the majority is more laborious than factor analysis as well as wasteful of information, for we learn nothing of the other factors and of the factor composition of those which fell by the wayside. Needless to say, most item analyses do not go to this conclusion. They stop after one, or possibly two, purgings—which is no stopping point at all, except that it has gotten rid of a few really small minority items—for the internal criterion is in a process of transition from an accidental starting point to a still remote end position of stability and factorial homogeneity.

Loevinger (**86**) has argued that collections of items should first be item-analyzed for this internal consistency and that various separate collections of items should then be used as unitary variables for a factorization. The process is sometimes recommendable, but the escape which this offers from a frontal attack by factorization is dubious. When one knows the mixed factorial composition of each item-analyzed collection, the problem of finding a means of purifying it is still complex. A better approach would seem to be to factorize a collection of *items*, each item being a variable. Then each subgroup of items highly loaded in one factor can be used as a criterion about which to collect, by item analysis, from a large number of promising items, those which actually measure this factor. Such an approach has recently been used in building the *Sixteen Personality Factor Questionnaire* (see **30**). This conclusion would make the factorization of relatively large matrices a more necessary, important, and prevalent problem than it has so far seemed to be.

(1), it is possible to take *only variables which have turned out to be clear markers for the factors in the first matrix.*

Methods (2), (3), and, especially, (4) are preferable if one can afford to wait to analyze the matrices in sequence; and (4) is preferable to (2) because estimates of factors take rather a lot of time and, if based on groups of variables which overlap for some of the factors, almost invariably lead to spuriously high interfactor correlations.

The choice of population for efficient factorization has perhaps already been sufficiently discussed in Chapter 19. The number of subjects should not strictly be regarded as limiting the number of factors one can take out—*all* factors must in any case be taken out if rotation is to be unbiased—but it *does* limit the accuracy of the loadings. The expression for the significance of the pattern of loadings on a single rotated factor (which is what one ultimately has to consider) has already been given (page 304), and it is the working out of this in relation to the situation presented by a given design which must decide the minimum size of population we can safely use. When many factors are anticipated, the *mean* loading on one is likely to be small, so a larger population is necessary. The magnitude of the uniqueness in a variable is also inversely related to factor significance, so a larger population is necessary when it is suspected that the variance of common factors is going to be small. When factor number and variable specificity are held constant, the significance of loadings (or at least the size of χ^2) increases as a linear function of the size of population.

Some rough ideas of the practicable minima of population size may also be gained empirically from a survey of existing successful researches; though it must be kept in mind that almost all have taken the smallest possible population for the results achieved, and that some purposes of factorization—the exploration of factor outlines in new fields—can proceed on smaller populations than are needed for further objectives. While one might reasonably aim to define the principal factors operating in a certain realm on as few as twenty variables and eighty persons, it is desirable for other purposes, e.g., factor estimation and individual prediction, to have about fifty variables and four or five hundred subjects as a minimum. For a first exploration of the factor patterns, however, it is obvious that using the latter number of subjects would unnecessarily increase the

labor of computing correlations, which is a substantial part of any factorization. It would be more convincing, if time for five hundred subjects exists, to make two separate factorizations on, say, 250 subjects each, since the rotation for simple structure is always a process open to some doubt, and two independent discoveries of the same simple structure would add greatly to the effectiveness of the total research.

If fewer than forty subjects[2] exist, it would in general be better to choose the experimental design of Q-technique, increasing the number of variables to a hundred or more. As indicated earlier (page 98), there cannot be a real *saving* from Q-technique, since tests must be multiplied when persons are reduced in number and the total testing time and correlating time then remain approximately the same as before. Neither can one generalize that the factor structure found in quite small samples will hold for larger, more normal samples. The initial choice of P-technique, on the other hand, may constitute a real saving, for the (individual) testing time is reduced for the experimenter (though much increased for the subject) even though the computing time is not altered.

CHOICE OF A CORRELATION FORMULA

The second area where labor may be saved or wasted—that of correlation computation—will be explored on the assumption that the size of population of persons and variables is held constant. Then the alternatives are to choose various correlation coefficient formulas and various computation aids. As pointed out in the previous chapter, a device such as the tetrachoric coefficient or the phi coefficient, in which a distribution of scores is cut at some arbitrary point such that all above are considered as one category and all below as of another, necessarily loses some of the available information. Consequently, for a population of the same size, it has a larger probable error. For this reason and because of doubt regarding the elimination of eccentricity factors (page 324), one would therefore normally prefer the product-moment formula. But the tetrachoric lends itself to rapid calculation by I.B.M. cards, being perhaps ten times as rapid as

[2] The numbers discussed here refer to the situation in which the total available population from which the sample is taken is for practical purposes infinite. Where, as in a population of nations, the total population is itself limited (about 90 in this case), forty cases, especially if chosen as a stratified sample, would not be insufficient for R-technique.

product moments calculated by ordinary computing machines or two or three times as rapidly as by I.B.M. Moreover, it uses less complex I.B.M. equipment than does the product moment, for it requires the sorting machine only and can be learned in ten minutes[3] by a computing clerk. Consequently there may be a saving of some magnitude in offsetting the larger standard error by using a larger population of subjects and employing the tetrachoric. Diagrams for reading off the coefficient from the fourfold percentage table are available by Chesire et al. (**35**), Hamilton (**62**) and others. The risk of sign errors in this process is considerable, a normally careful clerk making as many as 5%; so checking of signs by a different computer is absolutely essential.

[3] The procedure for working I.B.M. tetrachorics on a sorter has to be designed to give the least trouble in the subsequent reading off of coefficients from graphs, using the fourfold table, and experience shows that it is worth taking pains to avoid dangers of sign errors resulting from awkward transpositions having to be made in this latter stage.

Accordingly, it is best to begin by punching each individual's card with a hole at the appropriate position when the individual is *above* the agreed point of cut on that variable, leaving the absence of a hole to indicate that he is *below*. (Incidentally, this economy of representation permits many variables—more than one is ever likely to tackle in one factorization—to be represented on one card.) However, the direction of positive or above average score on each variable must be *arbitrarily decided* (as far as psychological meaning is concerned) *before* the above punching, on the principle that the *smaller fraction of the population is considered to have a positive score*. This is done for convenience in later reading from graphs. For statistical purposes (accurate tetrachorics) one attempts to keep the fractions above and below the cut as near halves as possible, but the natural degree of difficulty of the item almost invariably makes the yes's fewer than the no's, or vice versa. The directions decided upon at this point should be securely recorded, preferably by altering the label of the variable on the chart or correlation matrix at once to be consonant with the direction taken as positive.

One then puts the cards through the sorter for variable 1, counting and setting aside the cards which are positive (punched). The number is put at the top of the matrix opposite the column and row for variable 1. Then one takes this pack of positive cards and counts them with respect to all the other variables in turn, picking out the number positive for each. It is not necessary actually to sort the cards, and indeed the machine can be set to count up for as many as twelve of the associated variables at one pass. The figure for each of the associated variables—the number positive on both variable 1 and the other variable—is entered in the cell in the matrix under variable 1 and opposite the other variable.

To transform these numbers to tetrachorics in a corresponding matrix requires (a) first transforming the numbers to percentages of the total population, which is best done by transferral to an interim matrix, and (b) entering a graph with these percentages to get the tetrachoric coefficient, as described elsewhere (**35**, **62**).

In spite of these savings by the tetrachoric, its assumption of normal distribution and its loss of part of the available information will incline the careful worker to carry out really important researches (especially where the experimental data does not permit one to adjust the cut for the tetrachoric to be near the middle of the range) in terms of product-moment correlation. With proper equipment, indeed, the speed of product-moment calculations for large matrices can be made of the same order as for tetrachorics. Even before turning to devices using less common equipment, it is possible, first, to keep alert to the timesaving devices commonly taught in elementary statistics, notably that of using the *raw* score formula[4] (which follows the usual rule that an algebraically complex-looking formula generally means simpler arithmetic!) and employing nomographs.[5]

Secondly, one can use the device of rescaling all results to an approximately normal distribution with the same mean and standard deviation. This saves considerable time by giving one and the *same denominator* to all the hundreds or thousands of r's, and by making *the calculation of the numerator* a simple process for an ordinary adding machine. It also clarifies the simple structure resolution through bringing score distributions to normality.

The method requires first that all the various scores in their various interactive or raw units be scaled to the same number of basic derived units by grouping. In order to avoid need for any substantial use

[4]
$$r_{xy}=\frac{N\Sigma xy-\Sigma x\Sigma y}{\sqrt{[N\Sigma x^2-(\Sigma x)^2][N\Sigma y^2-(\Sigma y)^2]}}$$
where x and y are the raw scores, and N is the size of population.

[5] Though most computers are familiar with the following device for speeding correlation by ordinary machines, it may justify brief mention. Its aim is to permit the simultaneous accumulation of Σx, Σy, Σx^2, Σy^2, and Σxy from a single operation. For this purpose, x and y values (probably two-figure numbers) are inserted widely apart in the keyboard as multiplicand, thus x 0 0 0 0 0 y. They are also inserted as x 0 0 0 0 y in the multiplier keyboard. On multiplying they behave essentially as $(x+y)(x+y)=x^2+2xy+y^2$, and the answers appear in the upper keyboard thus, separated by zeros. At the end of N such multiplications we have Σx^2, $2\Sigma xy$, and Σy^2 in the upper dial, and Σx and Σy in the lower. Naturally, precautions must be taken that the numbers do not run together, by breaking the series into parts if N is very large.

When these values are obtained once for each variable, multiplication with other variables can be hastened by putting as many as three or four in the keyboard, e.g., x multiplies u, w, and z, giving on one keyboard Σxu, Σxw, and Σxz. The Σu, Σu^2, etc. have meanwhile been obtained when the u column of the correlation matrix was started, i.e., when it played the role of x in the first paragraph, and similarly for w, z, etc.

of Shepard's correction, the new units should not fall much short of 14, while for convenience of calculation (and sometimes out of regard for the crudity of the raw scores) they should not much exceed 10 or 12. If we adopt 12, we shall next find how many of the population would fall into each of 12 equal intervals if the whole were in a perfectly normal distribution. For 2^{11} cases, i.e., 2048, the numbers at 12 equal intervals would run

Equal units	1	2	3	4	5	6	7	8	9	10	11	12
Frequency of cases	1	11	55	165	330	462	462	330	165	55	11	1

For smaller populations, say 500 and 200, respectively, the following approximations might be taken to give whole numbers near the binomial expansion (for though the distribution must be symmetrical, it need not be exactly normal).[6]

Step 1	2	3	4	5	6	7	8	9	10	11	12
500 cases 1	2	12	40	80	115	115	80	40	12	2	1

Step 1	2	3	4	5	6	7	8	9	10
200 cases 1	5	16	32	46	46	32	16	5	1

Taking the scores in rank order for each variable, one now rescales, giving a score of 12 to the highest, 11 to the next 2, 10 to the next 12, 9 to the next 40, and so on. This takes time, but it is well repaid, at least with a large matrix, for (a) the $\Sigma x_1 x_2$ can be worked out with an adding machine alone, the multiplications being largely of one figure numbers and none above the twelve times table, which

[6] Other divisions can be obtained from tables of normal distributions or the Kelley-Wood tables relating percentiles to standard scores. For most objectives of factor analysis there is little need to be concerned if, through having a small population (say, 100 cases), one has to use a bigger sigma (or a platykurtic curve of distribution) in order to spread out so few cases over a 12-point distribution. For a normal distribution is not essential and even the measures of factor significance are little affected by this use of distribution which, though flattened, is still symmetrical. Some psychometrists even slice up their distributions very simply (in equal blocks) according to a rectangular distribution, but it seems wiser to keep as near to a normal distribution as the number of cases will permit.

can be done mentally (or many at a time with a multiplying machine), and (b) the denominator is the same for *all* correlations. This rescaling to a ten- or twelve-point normal distribution is thus an ideal device with *many variables*, not too *many persons* (for ranking takes time), and especially when one has limited computing machinery.

The rescaling plan, as first described above, supposes the use of the raw score formula (page 326). A slightly quicker alternative, which, however, requires attention to signs in multiplying, is to use an odd number of groupings in the rescaled units, so that the mean score falls exactly on the middle value, and to assign scores directly *as deviations*, positive or negative from this mean. Thus a classifying of raw scores into eleven equal intervals would give a range from $+5$ to -5. The numerator of r (as $\frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$) is now simply the sum of these products (never greater than 5×5) put down directly from the rescaled columns on to the adding machine, and the denominator is the sum of squares of one column, quickly worked out once and for all.

CALCULATING CORRELATIONS BY I.B.M.

Third (except for some rather specialized graphical methods which may be sought in regular statistics books), the correlations may be worked by I.B.M., which is the quickest of all. Detailed description of I.B.M. methods of correlating (let alone of factor analysis proper) is not attempted here since there is so much variation in the machines that may be available. However, the basic process uses the principle of progressive digiting as a means of finding sums of squares, cross-products, etc. on the tabulating machines—which can only add or subtract. Thus, to *multiply* one number by another, we may add the first number for as many times as the second requires—thus:

$$6 \times 3 = 6 + 6 + 6 = 18.$$

When many cross-products are to be summed, certain subtotals are conveniently found first. Thus:

$$\underset{x_1\ \ y_1}{(4 \times 8)} + \underset{x_2\ \ y_2}{(3 \times 6)} + \underset{x_3\ \ y_3}{(2 \times 3)} = \underset{4}{(8)} + \underset{3}{(8+6)} + \underset{2}{(8+6+3)} + \underset{1}{(8+6+3)}$$

It will be seen that when these subtotals are properly ordered, each is equal to or greater than its predecessor. In fact, the nth subtotal

from the end is equal to its predecessor plus the values of y which are to be multiplied by n. Therefore, if cards bearing numbers for the various products to be formed are sorted according to the value of x, from high to low, all these subtotals will be formed in the course of one run of these cards through the tabulator, and the machine can be controlled to print these subtotals, and better still, punch new cards containing them. The new cards (called summary cards) are then run back through the tabulator causing it to find the final answer.

Actually, it is possible simultaneously to total for a whole group of y's, say y_a, y_b, etc., which are to be multiplied by the same set of values for x, thus finding material for a whole series of r's at once. If y is set equal to x, then the sum of squares will be found.

The various computations carried out must all be completely identified in any manner appropriate to the machine steps being utilized. Then, at the end of all the summing, the various sums belonging to one correlation are brought together into one card. These cards may then be used to print lists of Σx, Σy, Σx^2, Σxy, Σy^2 for hand computation of the r's, or, if an I.B.M. multiplier is available, it may be wired to compute r^2 from the sum.

Granted some such radical abbreviation of the correlation computations, the three major aspects of factorization—correlation, factor extraction, and rotations—make about the same demands on time when the normal timesaving steps are taken in all processes, as indicated below. But otherwise, a major part of the total time available must be allotted to correlation.

ECONOMY IN EXTRACTION OF FACTORS

The devices for shortening the next step—factor extraction—have already been fully discussed in Chapters 10 and 11, so that only the briefest summary is necessary. Principal component and maximum likelihood methods are the longest, though the former is likely to prove a quick and convenient method when electronic computers are available. The basic centroid method is intermediate. The group centroid methods are fast, and of these the fastest is the multigroup method, which can advantageously be made the standard method where highly skilled assistance is available. But for lower levels of training in the computing laboratory the *grouping* method almost certainly offers the best compromise on quickness, adequate checking devices, and accuracy.

Burt has claimed that the quickest of methods is his shortened

method by submatrices (**10**) and, indeed, there is no doubt that this method and its close relatives—Holzinger's bifactor (**71**) and Woodrow and Wilson's (**143**) suppression method can beat most centroid methods in computing speed itself. The objections to them are (1) that a lot of stage setting has to be done arranging in submatrices before the actual computation begins; (2) that in the Burt and bifactor methods the process is complicated when there are many negative correlations; (3) that in these devices there is more rotation to be done later in finding simple structure than with the grouping centroid methods; and (4) that many matrices simply do not yield relatively independent clusters of the kind required.

Woodrow's method, still little tried, perhaps has most promise of these methods. It comes out nearer to a rotated position, but with small matrices it is not easy to reflect variables to get the suppression of unwanted variables (with respect to a given group) which this method requires in the noncluster variables. Probably these methods, at least that of Woodrow, are best considered as instruments to be kept available for specially favorable occasions—positive manifolds and very well-defined, independent clusters—and especially where a quick, approximate answer is required. There are, in fact, several such approximate, relatively direct methods of proceeding to a *rough* answer if the experimenter is alert to special cases and knows his factor analytic principles well. For example, there is Tryon's cluster analysis (**128**) variously modified by Osgood and others. Still more useful is the trick of looking out for two or three variables (actually, as many variables as one suspects there are factors) which happen to have zero correlations with one another. These can be taken directly as factors, the correlations of the other variables with them being the loadings of the latter. Such uncorrelated variables can actually be inserted in the initial battery, but one must be sure they will remain uncorrelated in the given sample and that they will have substantial correlations with the other variables.

A recent suggestion for a quick and approximate solution is that of Mosier (**152**), which picks out groups, preferably according to prior knowledge of factors, in an explored field. The intercorrelations of variables in the group are computed, and the matrix is tested for rank one. If this is approximately sound, the correlations of all other variables in the matrix are computed with the score on this set, giving their "loadings" in that factor, and so on for any other reasonably independent clusters. Here one only does a *factor analysis*—to the

extent of testing for rank one—on the restricted group in one cluster, but the remaining calculation involves using the scores on the set, with the aid of a score pattern key, which can be time-consuming. Like most approximate methods it is really useful only where factors are very few, and can be used with confidence best in fields of variables already once explored by regular methods.

But most approximate devices of this kind can be used only for rough descriptive and exploratory purposes, not for erection upon them of specification equation predictions; and even for describing the nature of factors, they are likely to cause confusion where factor patterns are only distinguished by subtle differences.

With the main multigroup method it is often convenient to take out first one large batch of factors, but not so many that any difficulty is likely to be presented in finding at least one further cluster. One assistant can then begin rotating these factors while the other feels his way more tentatively to the extraction of the last one or two factors and the rejection of factors that have imaginary numbers and are therefore in excess.

SHORTENING THE MULTIGROUP MATRIX INVERSE CALCULATION

In an earlier chapter (page 224) it has been pointed out that one of the most difficult, or, at least, tricky and tedious processes that the factor analytic computer has to face is the calculation of inverses of matrices. Inverse computations are routinely encountered (1) in the comparatively brief and simple calculation of the inverse of a *triangular* matrix when the multigroup extraction method is used and (2) in its complete complexity when we wish to transform correlations on the RV's to loadings on factors. Two methods for routinely computing the latter have been given on page 226. An immense saving is possible for those to whom an electronic digital computer is available; for the usual instrument of this kind can calculate the inverse of a matrix as large as 30×30 in about five minutes.

As for the computing of a triangular matrix, a method due to Saunders is already set out on page 182. Essentially the same design with some differences of procedure has been published by Fruchter (**56**), and since it develops the argument in more detail and with illustrations, it is reproduced here, by kind permission of its author, as an alternative to meet the fuller interests in computing processes of those reading this chapter.

Let us start with a matrix giving the mutual angles of the temporary factors f_1, f_2, f_3, etc. obtained from the multigroup extraction process. For example, in a three-factor problem, taken for illustration from Thurstone (**126**), we might have:

	f_1	f_2	f_3	
f_1	1.000	.467	.342	
f_2	.467	1.000	.437	$=C_{ff}$.
f_3	.342	.437	1.000	

We first factorize this by the diagonal method, obtaining the λ matrix of which this C matrix is as usual the product of λ and its transpose ($\lambda'\lambda = C$). Thus we obtain:

	F_1	F_2	F_3	
f_1	1.000	0	0	
f_2	.467	.884	0	$=\lambda_{fF}$
f_3	.342	.314	.886	

giving the angles of the f's to the required unrotated factors, the first of which is arbitrarily placed along the first of these multigroup factors.

It is the inverse of the transpose of this which is required to calculate the F loadings from the f loadings. Let us represent the above particular case by a general case, which we shall call λr.

	F_1	F_2	F_3	
f_1	1.0[1]	0	0	
f_2	r_{21}	r_{22}	0	$=\lambda r$
f_3	r_{31}	r_{32}	r_{33}	

[1] The value in cell r_{11} (here f_1F_1) is always 1.0 for this type of problem.

Represent the inverse of this matrix $(\lambda fF)^{-1}$ by λr^{-1}

	F_1	F_2	F_3
f_1	c_{11}	c_{12}	c_{13}
f_2	c_{21}	c_{22}	c_{23}
f_3	c_{31}	c_{32}	c_{33}

(The values in the lower left will actually be zeros.) Then $\lambda r \cdot \lambda r^{-1} = F_1$, or, written out:

$$\begin{array}{c|c|c|c} & F_1 & F_2 & F_3 \\ \hline f_1 & 1.0 & 0.0 & 0.0 \\ f_2 & r_{21} & r_{22} & 0.0 \\ f_3 & r_{31} & r_{32} & r_{33} \end{array} \cdot \begin{array}{c|c|c} F_1 & F_2 & F_3 \\ \hline c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{array} = \begin{array}{c|c|c} F_1 & F_2 & F_3 \\ \hline 1.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 1.0 \end{array} \tag{50$_1$}$$

Performing the row by column matrix multiplications gives the following equations:

(Row 1×column 1)

$$1.0c_{11}+0.0c_{21}+0.0c_{31}=1.0 \tag{50$_2$}$$

$$c_{11}=\frac{1.0}{1.0}=1.0 \tag{50$_3$}$$

(Row 2×column 1)

$$r_{21}c_{11}+r_{22}c_{21}+0.0c_{31}=0 \tag{50$_4$}$$

As shown in equation (3), $c_{11}=1.0$. Hence

$$r_{21}+r_{22}c_{21}=0 \tag{50$_5$}$$

$$c_{21}=-\frac{r_{21}}{r_{22}}=-\frac{0.467}{0.884}=-0.529 \tag{50$_6$}$$

(Row 3×column 1)

$$r_{31}c_{11}+r_{32}c_{21}+r_{33}c_{31}+0 \tag{50$_7$}$$

From equation (3) $c_{11}=1.0$, and from equation (6)

$$c_{21}=-\frac{r_{21}}{r_{22}}$$

Substituting and transposing

$$r_{33}c_{31}=-r_{31}+\frac{r_{32}r_{21}}{r_{22}} \tag{50$_8$}$$

and

$$c_{31}=-\frac{r_{31}}{r_{33}}+\frac{r_{32}r_{21}}{r_{22}r_{33}}=$$

$$-\frac{0.342}{0.886}+\frac{0.314\times 0.467}{0.884\times 0.886}=-0.386+0.187=-0.199 \tag{50$_9$}$$

(Row 1×column 2)

$$c_{12}=0.0 \tag{50$_{10}$}$$

(Row 2×column 2)

$$r_{21}c_{12}+r_{22}c_{22}+0.0c_{32}=1.0 \tag{50$_{11}$}$$

From equation (10) $c_{12}=0$; hence

$$c_{22}=\frac{1}{r_{22}}=\frac{1.0}{0.884}=1.131 \tag{50$_{12}$}$$

(Row 3×column 2)

$$r_{31}c_{12}+r_{32}c_{22}+r_{33}c_{32}=0 \tag{50$_{13}$}$$

From equation (10) $c_{12}=0$, and from equation (12)

$$c_{22}=\frac{1}{r_{22}}$$

Substituting

$$\frac{r_{32}}{r_{22}}+r_{33}c_{32}=0 \tag{50$_{14}$}$$

$$c_{32}=-\frac{r_{32}}{r_{22}r_{33}}=-\frac{0.314}{0.884\times 0.886}=-0.401 \tag{50$_{15}$}$$

(Row 1×column 3)

$$c_{13}=0.0 \tag{50$_{16}$}$$

(Row 2×column 3)

$$c_{23}=0.0 \tag{50$_{17}$}$$

(Row 3×column 3)

$$r_{31}c_{13}+r_{32}c_{23}+r_{33}c_{33}=1.0 \tag{50$_{18}$}$$

From equations (16) and (17), $c_{13}=0$, and $c_{23}=0$. Therefore

$$c_{33}=\frac{1}{r_{33}}=\frac{1}{0.886}=1.129 \tag{50$_{19}$}$$

Putting the results in the form of a matrix gives

	F_1	F_2	F_3
f_1	1.000	.000	.000
f_2	−.529	1.131	.000
f_3	−.199	−.401	1.129

$=(C_{fF})^{-1}$

This is the inverse of C_{fF}. Comparing matrices C_{fF} and $(C_{fF})^{-1}$, it may be observed that wherever a zero occurs above the principal diagonal in the former it also occurs in the latter. The values along the principal diagonal of $(C_{fF})^{-1}$ are the reciprocals of the corresponding values of C_{fF}. The other values of $(C_{fF})^{-1}$ are obtained by means of simple equations similar to those outlined above.

The desired matrix for multiplying the f values to get the F loadings is therefore merely the transpose of the above and is written

$(C'_{fF})^{-1}=$		F_1	F_2	F_3
	f_1	1.000	−.529	−.199
	f_2	.000	1.131	−.401
	f_3	.000	.000	1.129

This is the required inverse of the first diagonal matrix and it can now be applied at the end of the multigroup extraction to get that original, unrotated, orthogonal V_0 from which the rotation process may best begin.

EXTENDING A MATRIX ALREADY FACTORIZED

A device for shortening factorization, which has already been briefly mentioned in connection with possibilities of dividing up a large battery into smaller matrices, is Dwyer's extension and other extension systems. They aim to break up the variables into a majority, which is factored as a single matrix, and a minority, the loadings of which are found by obtaining later their correlation with the first set. Two such methods are set out below, the first due to Dwyer (44) and the second to Saunders (private communication). The first is laborious and must be regarded as a convenience for special circumstances rather than a means of shortening factorization generally. Its convenience lies in the fact that one is sometimes ready to go ahead with factorization of the majority of variables but is held up while waiting for data on two or three more to come in, and this device permits one to factorize in the usual way and determine the factor loadings of the stragglers later (or of special tests applied to the group later as a result of some hypothesis reached by the factorization of the majority).

That it would theoretically be possible to extend the correlation matrix, i.e., to add columns representing the r's of new variables with the old, and obtain therefrom the factor loadings of new variables with the factors found among the old follows from the fact that we can estimate any factor in terms of the old variables. Naturally, there is a slight loss of accuracy in the extension. Consequently, it is desirable to get the more important variables into the main body

of the factorized matrix and the more exploratory ones into the extension.

The method can theoretically be used for any number of variables in the extension, and the following are the steps to be followed in the general case:

1. In the V_0 matrix, we let t equal the number of rows and k equal the number of columns. (That is, there were t tests and k factors were extracted.)

Call each factor loading a_{ij} where i is the row number, j is the column number of the loading. (Thus, the loading in the eighth row and third column would be a_{83}.)

2. For each new test to be added to the V_0 matrix, let its correlation with the t given tests be $r_1, r_2, r_3, \ldots r_t$. (If there were 30 tests in the original V_0 matrix, then for each new test to be added we must have 30 values of r.)

$V_0=$

Test Nos.	Factor numbers 1	2	......	k
1	a_{11}	a_{12}		a_{1k}
2	a_{21}	a_{22}		a_{2k}
3	a_{31}	a_{32}		a_{3k}
.				
.			(a_{ij})	
.				
.				
t	a_{t1}	a_{t2}		a_{tk}

$A=$

Factor Nos.	Factor numbers 1	2	......	k
1	A_{11}	A_{12}		A_{1k}
2	A_{21}	A_{22}		A_{2k}
.				
.				
.			(A_{ij})	
.				
.				
k	A_{k1}	A_{k2}		A_{kk}

$$\left.\begin{cases}[A_{21}=a_{11}a_{12}+a_{21}a_{22}+a_{31}a_{32}+\ldots+a_{t1}a_{t2}]\\ [A_{11}=a_{11}^2+a_{21}^2+a_{31}^2+\ldots+a_{t1}^2]\end{cases}\right\}\text{etc.}$$

3. Multiply each loading in column 1 of the V_0 matrix by its corresponding loading in column 2 ($a_{11}\times a_{12}$, $a_{21}\times a_{22}$, $a_{31}\times a_{32}$, . . . , $a_{t_1}\times a_{t_2}$) and find the sum.

In a new matrix, record this sum in the second row, first column. Repeat this process using column 1 with each of the other columns of the V_0 matrix, recording all the sums in the first column of the new matrix, which should then have as many rows as there were factors in the V_0 matrix (the first row is still blank). Now replace

column 1 by column 2 in the above instructions and fill in column 2 of the new matrix beginning with the product of columns 2 and 3 of V_0, which will be entered in the second column, third row of the new matrix.

Continue in this manner until each column of V_0 has been multiplied by every other column. The new matrix is now completely filled in below the diagonal. Recopy this lower half into the upper half as in the grouping methods of factor analysis.

In the main diagonal, enter the sums of the squares of the loadings in each column of V_0, i.e., the sum of the squares of the numbers in column 1 of V_0 is entered in the first row, first column of the new matrix; the sum of the squares of numbers in column 2 of V_0 is entered in row 2, column 2 of the new matrix; etc.

The new matrix is now complete and contains as many rows and columns as there were factors in V_0; that is, it is a square matrix of k rows and k columns. Call this matrix A and the element in its ith row, jth column, A_{ij}.

4. Find the value of the determinant of A (see **126,** page 5).

5. Find the value of the cofactor of each A_{ij} (see **126,** page 7).

6. Divide each number found in 5 by the number found in 4 and enter these values into another matrix D, similar to A. (In D, each element D_{ij} is the result of dividing the cofactor of A_{ij} by the number found in 4. The upper and lower halves of D will be reflections of each other, as were the two parts of A.) This is equivalent to finding the inverse of a matrix.

7. For each test to be added to V_0, find the sum of the product of its correlations with the original tests by the loading on each factor. Thus, we form the sum of the products, $a_{i_1} \times r_i$, where i takes on as many values as there are original tests. (This step is similar to 3 with the correlations replacing one of the columns of V_0.)

Make still another matrix R, similar to A and D, in which the sums just found are recorded. In general, R will not be square, but will have k rows and as many columns as there are new tests to add.

8. Multiply the first column of D by the first column of R and find the sum of these products. This is the factor loading of the first new test in factor 1. Multiply the second column of D by the first column of R, find the sum; this is the factor loading of the same test in factor 2. Continue using the first column of R with each column of D, thus

obtaining the k factor loadings of the first new test. Repeating this process with each column of R will give the factor loadings of each new test with each of the k factors.

$$
D=\begin{array}{c|c|c|c}
 & 1 & 2 & \cdots\cdots\quad k \\
\hline
1 & D_{11} & D_{12} & D_{1k} \\
2 & D_{21} & D_{22} & D_{2k} \\
3 & D_{31} & D_{32} & D_{3k} \\
\cdot & & & \\
\cdot & & & D_{ij}=\dfrac{(\text{cofactor of } A_{ij})}{A} \\
\cdot & & & \\
\cdot & & & \\
\cdot & & & \\
\cdot & & & \\
k & D_{k1} & D_{k2} & D_{kk}
\end{array}
\qquad
R=\begin{array}{c|c|c|c}
 & \multicolumn{3}{c}{\overbrace{\qquad\qquad\qquad\qquad}^{\text{Number of new tests to be added}}} \\
 & 1 & 2 & \cdots\cdots \\
\hline
1 & R_{11} & R_{21} & \\
2 & R_{21} & & \\
3 & & & \\
\cdot & & & \\
\cdot & & & \text{new factor} \\
\cdot & & & \text{loadings} \\
\cdot & & & \\
\cdot & & & \\
\cdot & & & \\
k & & &
\end{array}
$$

$$
\begin{array}{c}
\text{New} \\
\text{correlations} \\
\text{(old test} \\
\text{numbers)}
\end{array}
\left\{
\begin{array}{c|c|c|c|c|c}
 & \multicolumn{5}{c}{\overbrace{\qquad\qquad\qquad\qquad\qquad}^{\text{Number of new tests}}} \\
 & (1) & (2) & (3) & \cdots & \cdots \\
\hline
1 & r_1^{(1)} & r_1^{(2)} & & & \\
2 & r_2^{(1)} & r_2^{(2)} & & & \\
\cdot & & & & & \\
\cdot & & & & & \\
\cdot & & & & & \\
\cdot & & & & & \\
t & r_t^{(1)} & r_t^{(2)} & & &
\end{array}
\right.
$$

$$R_{11}=[r_1^{(1)}a_{11}+r_2^{(1)}a_{21}+\cdots+r_t^{(1)}a_{t1}] \text{ etc.}$$

$$R_{21}=[r_1^{(2)}a_{12}+r_2^{(2)}a_{22}+\cdots+r_t^{(2)}a_{t2}]$$

It may be recognized that the whole of the above process is based essentially on the equation $(RV_0)\ (V_0'V_0)^{-1}=\text{Extended } V_0$.

While the procedure just given for Dwyer's extension must be used when the factor analysis has been performed by any of the methods requiring the computation of successive residual matrices,

and avoids the computation of residual correlations involving the new variables, a shorter method is applicable if the multiple group centroid procedure has been followed for the original factorization. The computations are really the same, but most of the work required by Dwyer's extension has already been performed in this case as a part of the routine procedure for the multiple group centroid extraction. Three steps are then required:

1. The original correlation matrix is extended by writing the new correlations in columns at the right of the old columns. Each new variable must be correlated with each variable which was used in any group of the group centroid procedure, but no correlations need be found between the new variables nor between a new variable and an old variable which was not used in any group.

2. The t matrix is extended by totaling for each new variable the correlations it has with the members of each group, in turn. In other words, the columns added to the correlation matrix are treated in the same manner as they would have been in the original factor analysis, except that the new variables cannot become members of any group.

3. The V_0 matrix is extended by treating the addition rows of the t matrix in the same manner as though they had been there all along. In the same way, the rotated matrices are extended by treating the additional rows of the V_0 matrix in the same manner as the old ones.

In a computing unit which habitually uses the multigroup extraction method, this method as set out by Saunders involves much less work than the Dwyer method, and can constitute a real saving of time over the normal extraction process. However, it should be noted that the extension variables are assessed on the centroids of the original matrix variables. The extension methods cannot therefore yield any additional factors that might be (a) normally found among the extension variables but not the original contingent, and (b) arising between the extension and the original, but not found in the latter beforehand.

FACTOR EXTRACTION BY I.B.M.

The most radical aid at present possible in the factor extraction process is the use of I.B.M. machine methods. It is not worth while to attempt this, however, with small matrices or without the coöperation of an I.B.M. technician, for it is still more complicated than

the procedure outlined above for obtaining correlations by I.B.M. The procedure will depend on the particular machine available.

An early design for I.B.M. centroid factorization was put forward by Tucker, extended by Hall, Welker, and Crawford (**61**), and improved further by Saunders and Hall. Tucker has also described in detail an I.B.M. card method for a principal components factor extraction (**132**). The following account of the Saunders-Hall procedure for *centroid* extraction assumes an I.B.M. unit with only a key punch, sorter, and tabulator, though it proceeds more quickly with a summary punch (the absence of which will multiply the work by three or four). The reader will understand that detailed description is difficult in view of the diversities in machines available. However, the elementary steps involved in (a) the finding of (integrally) weighted sums of correlations, and (b) the finding of products and residual matrices can be translated into the terminology of the I.B.M. technician, and one or two of the more troublesome problems of detailed arrangement can be considered.

First, the general issues. Correlations and residuals are usually given to six figures including four decimal places. Negative correlations and residuals are customarily carried in complement form in order to minimize the problems of sign combination and summary X-punching. Communalities of 1.0000 are punched with the original correlation matrix, and communality correction cards are prepared afresh for each iteration if reëstimation is to be indulged in. From five to twelve correlations will be punched in each card, depending on the counter capacity of the tabulating machinery available. Each card will be identified according to the row of the matrix, and the first column from which the correlations are taken (every correlation for a given column must be in a different card), as well as the number of the iteration, the function of the card, and the designation of the study. It is convenient to designate by a letter the groups of from five to twelve variables, and to number the variables within each group; in this way only three columns are required, and the same designations identify the same variable whether it appears in a row or a column.

Certain auxiliary decks of cards are necessary. First, a set of subtraction setup and clear cards is required. These must be identified so as to sort along with the column group identifications in the main deck. One setup and one clear card is required for each group.

Second, a multiplying deck is required, which is punched with the last digit of the column number for the first seventy columns and contains extra sources of zeros as well as identification in the set for the remaining ten columns. The requirements of these decks will be apparent from the mode of their subsequent use.

The subtraction setup and clear cards are separated into (a) the setups and (b) the clears. The cards of the correlation matrix are then sorted by row group and row, keeping the individual rows separate. Rows to be given a weight of plus one are filed in front of the subtraction setup cards, rows to be given a weight of zero are set aside, and rows to be given a weight of minus one are filed between the subtraction setup and clear cards. When all the variables have been assigned weights, the deck is sorted by column group and run through the tabulator with control on the column designation. The weighted totals are printed and may also be summary punched if subsequent trials are to include the total of the present one. The process of sorting and reweighting may be repeated any number of times, according to the general method of factorization which is being employed. If the centroid method is being followed, communality suppression cards will be included along with the correlation matrix deck until the proper signs have been determined, at which time further cards bearing the estimated communalities may be added to the deck for a final run.

When suitable column totals have been obtained, by whatever sequence of operation, the loadings on the factor are computed according to Guttman's formula

$$a_{jk}=\frac{\Sigma_i\omega_{jk}r_{ij}}{\sqrt{\Sigma_j(\omega_{jk}\Sigma_i\omega_{ik}r_{ij})}} \tag{51}$$

to two decimal places. I.B.M. equipment is not used for this step. A new deck of cards is then prepared, with each card containing one variables row identification and its loading (with negatives as complements). This deck is sorted with the multiplying deck on the columns containing the loading values of the former and the identification numbers of the latter. A multiplying board is wired for the tabulator utilizing the main brushes as a digit emitter so that the loadings of a column group are added once for each card of the multiplying deck. By progressive digiting, all the possible products of these by loadings are set up successively in the counters of the tabu-

lator; and the loading cards sorted into the multiplying deck control the punching of appropriate summary cards containing these products. The multiplying board is rewired from the main brushes for each column group until all the cards of the product matrix have been obtained.

The cards of the product matrix are then sorted together with the original correlations and run through the tabulator with control on each pair of cards; the summary punch produces a deck of residual correlations which may be used to start the cycle of operations at the beginning again.

The I.B.M. multiplier may be used for the computation of (non-integrally) weighted sums of correlations, such as required for the determination of exact principal components or the maximum likelihood factors of Lawley, but such procedures do not appear to have been worked out as yet.

Only minor modifications of the above are required for its use with the multigroup extraction process.

SUMMARY OF ROTATION ECONOMY

Methods for greatest economy in rotation have been adequately discussed individually, and it was pointed out that without a detailed cost accounting for different kinds of problems, no adequate generalization about their relative quickness could be made. However, when the study in question is running over familiar ground, a very substantial saving of time is certain if one proceeds by (a) moving at once to trial reference vectors which have the mean direction cosines of the half dozen variables known to be quite highly loaded in each of the factors already well recognized in the field; and (b) taking the group of variables which now appear to lie approximately in the hyperplanes of these factors and using the most rapid analytical method (page 274) to pull the reference vector perpendicular to whatever is the true form of this hyperplane.

In the majority of studies made in this decade, however, it would be a presumption and a serious danger of perpetuating error to rotate to trial vector positions chosen by supposed previous knowledge of factor resolution. The basic sectional view method carried out blindly in the sense of ignoring one's prejudices about the variables and shifting single reference vectors is then the only sure way of proceed-

ing; and it is by finding ways of speeding this process that the greatest contributions can be made in general to the economy of rotation.

Considerable time saving results—perhaps to the extent of reducing the usual process to a third of the time commonly required—by a combination of relatively simple devices, as follows: (a) Use graphical intermittent runs (page 264) of about three to six successive rotations according to the accuracy of the individual and his instruments and to the stage of the rotation (the last steps need to be more accurate). (b) Save time in drawing graphs by working only with even- or odd-numbered variables when the total is forty or more. (Switch from even to odd at the time of recalculating the V_n matrix at the end of a run.) (c) Use only a half or two-thirds of the total possible number of graphs, by stopping drawings on a given factor as soon as a good shift appears, and by choosing drawings where (i) a poor inter-*R.V.* angle needs to be straightened or (ii) there are many variables in common in the hyperplane (page 262). (d) Work with skeleton matrices in which the values in V_0, λ, and C are kept rounded to one decimal place. This speeds the calculation processes and introduces surprisingly little error in the rotation.

These abbreviated processes can be followed practically to the last rotation. Indeed, better results are obtained in a somewhat shorter time by making, say, ten general rotations, the first nine of which employ the above approximations, than by making five carried out with all possible accuracy at each stage. In the last resort economy in rotation results from having it done by a person with a gift for spatial thinking and a judgment not easily led astray by inessentials, who can judge, for example, when to take a hyperplane as wide and when to consider it narrow, when to straighten an angle and when to leave it, without any rigid mechanical following of rules.

However, the greatest scope for economizing on rotation, especially by the trial-and-error sectional view method, lies in the improvement of machine aids. The first device here—and as yet the only one with which experience has accumulated—is the matrix multiplier as adapted from the I.B.M. scoring machine by Tucker (**129**) and mentioned above (page 260). A whole column of the λ matrix is marked on a score card and put in the machine. The computer puts in by a keyboard the successive rows of V_o and gets at each press of the button an entry for a value in the given column of V_n. The answer tends to be about as rough as if one had used values rounded to one decimal place.

But a great advance in speed in rotation processes is available now through the I.B.M. multiplier (a rotation of a 15×18 factor matrix can be done in a day). Possible greater advance (though coding time will reduce it) is promised by the use of the electronic digital calculators which can multiply out the product of two whole matrices in a few minutes. If the matrices are as large as frequently occurs in factor analytic work—say, twelve factors and eighty variables—it may be necessary, because of the restrictions on memory which still hamper these calculators, to split the matrix into two parts. As there are now nearly a dozen such calculators distributed about the country, it seems worth while to set out in the appendix of this book (page 431), for the general guidance of those fortunate enough to have access to them, the procedure by which this vast saving in matrix multiplication may be effected.

When correlation, extraction, and rotation are complete, there remains the setting out of results, in which, as indicated on page 223, a further saving may *generally* be made by leaving the findings in terms of correlations with the reference vectors instead of loadings on the factors. Although it is *always* a slight advantage to the reader to be given the latter, yet where the relative variance of different factors does not need to be exactly known, and where the specification equation is not needed for individual predictions, this additional labor of perfection should not be demanded of the researcher, though he should clearly, explicitly state that the results are presented in reference vector terms. Moreover, he should omit nothing that is required if the reader wishes to derive any transformations that may be needed to test hypotheses or plan further research (see page 232).

MECHANICAL AND MACHINE AIDS

Finally, in an effective computing laboratory, it will be found that at *all* stages of factor analysis one cannot despise the help to be gained from quite small mechanical aids and seemingly trivial regulations which insure that all those working on a project use the same devices and symbols to reduce errors and avoid misunderstandings when one person takes over from another. In most laboratories nowadays the efficiency of computing is greatly helped by proper attention to the process of communication and recording within small groups of co-workers. For example, a sheet should be prominently set up bearing the meaning of the chief symbols in use. Printed matrix forms with

	$b=-a$																						
	a	Test	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
		1																					
		2																					
		3																					
		4																					
		5																					
		6																					
		7																					
		8																					
		9																					
		10																					
		11																					
		12																					
		13																					
		14																					
		15																					
		16																					
		17																					
		18																					
		19																					
		20																					
		21																					
		22																					
		23																					
		24																					
		25																					
		26																					
		27																					
		28																					
		29																					
		30																					
		31																					
		32																					
		33																					
		34																					
		35																					
		36																					
		$=V=\Sigma a$																					
		$s+vb+b^2=Ch$																					
		$\Sigma r=s$																					
		$-1/2\,s=A$																					
		( )																					
		( )																					
		( )																					
		( )																					
		( )																					
		( )																					
		( )																					
		( )																					
		( )																					
		( )																					
		( )																					
		B ( )																					
		$-2B=C$																					
		d																					
		$\mu d=D$																					
		$C-D=E$																					
		$mE=a_j$																					

DIAGRAM 35. Matrix Sheet

															$V=$		
22	23	24	25	26	27	28	29	30	31	32	33	34	35	36		Σ	CHECK

TABLE NO.________

________________Matrix

Test Battery________________

Date:________________

Computer:________________

$= \Sigma^s - v^2 + \Sigma b^2 \div \Sigma Ch$ $\square$ $\div \Sigma\Sigma'$ $\square$

$T = \Sigma|E| =$ $\boxed{}$

$\sqrt{T} =$ $\boxed{}$

$1/\sqrt{T} = m =$ $\boxed{}$

$mT =$ $\boxed{}$ $= \sqrt{T}$ $\square$

$\Sigma|a| =$ $\boxed{}$ $= mT$ $\square$

$= m\Sigma E = \Sigma a$ $\square$

for Extracting Factors.

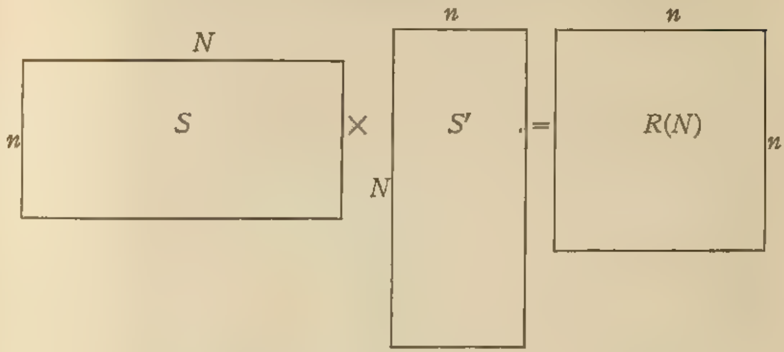
rows and columns of standard width help rapid transfer of data. In some processes the reading of figures onto a wire recorder and their repetition therefrom is quicker than writing out. Also there are such helpful trivia as marking groups in the grouping methods of extraction by red pencil rails above and below the columns and rows concerned; putting totals first in pencil and changing to ink only when checked; omitting positive signs and decimal points in matrices; writing in the negatives by a long dash; keeping graphs on the same scale and with the lower (or alternatively, the even) numbered factor vertical; copying out V_0 (unrotated) matrices in duplicate—for safety and to enable two computers to calculate V_n matrices simultaneously; checking the V_0 matrix as soon as obtained against the correlation matrix by taking two or three dozen inner products; keeping the width of columns in λ and V matrices the same for ease of multiplication, etc.; using standard printed matrix forms of some such kind[7] as shown in Diagram 35; running Scotch tape over the folding edge of much-used matrices to save figures and paper from abrasion; and labeling all forms clearly, e.g., as Residual no. 2, Product no. 4, Correlation Matrix for study X, etc.

The key to any further substantial advance in economy of factor analytic processes—whether of extraction, rotation, or checking—lies in the development of electronic computing machines for matrix multiplication, the calculation of inverses, etc. Although some years must elapse before such help becomes widely available, and present lack of experience would make discussion premature, it is appropriate to conclude by reminding the reader of the essential matrix form of the main factor analytic process, making practically all steps adaptable to matrix multiplying devices. The condensed statement in this form will at the same time provide the best possible summary of the ground that has been covered. In the summary, n as usual will represent the number of variables, N the number of subjects, and k the number of factors.

First we may recognize that the calculation of the correlation coefficients themselves can be set up as a matrix multiplication. If the scores are first put in terms of standard scores, then the score matrix S, postmultiplied by its transpose S', gives the correlation matrix R, provided every term in the product is divided by N. This may be put in more easily visualized form as follows:

[7] Copies of this form (full size) may be obtained in quantity from the Institute for Personality and Ability Testing, 1608 Coronado Drive, Champaign, Illinois.

$$S \times S' = R(N) \tag{52}$$



(where n is the number of variables)

In contemplating this statement it is appropriate to indicate a recent development—direct factor analysis—which, while it is too laborious to supplant standard methods yet, has a theoretical neatness and completeness as well as a freedom from some restrictive statistical assumptions in the analysis of correlation matrices. The direct factor analysis method, as described by Saunders (**110**), factorizes the score matrix (S above) instead of the correlation matrix. As some of the equations below remind us, the result of R-technique factorization is a test factor matrix V_0, or (rotated), V_{Fn}, expressing the loadings of the tests in the factors. From this and the score matrix we can estimate each individual's degree of possession of each factor which can be expressed in a population-factor matrix P; or we can obtain this matrix by Q-technique.

Now the relation of the test factor matrix and the population factor matrix to the score matrix is as follows:

$$S = (V_{Fn}\lambda^{-1})(\lambda P) \tag{53}$$

where λ is the nonsingular transformation matrix used for transforming V_0 into V_{Fn}, i.e., to express the process of rotation. Direct factor analysis aims to get these two matrices V_{Fn} and P from the score matrix, thus giving an R- and a Q-technique solution at the same time (and incidentally solving the problem of Q-technique rotation by tying it to the R-technique matrix which gives simple structure).

However, there is no saving in computation. The procedure consists in multiplying the score matrix by its transpose in a series of oper-

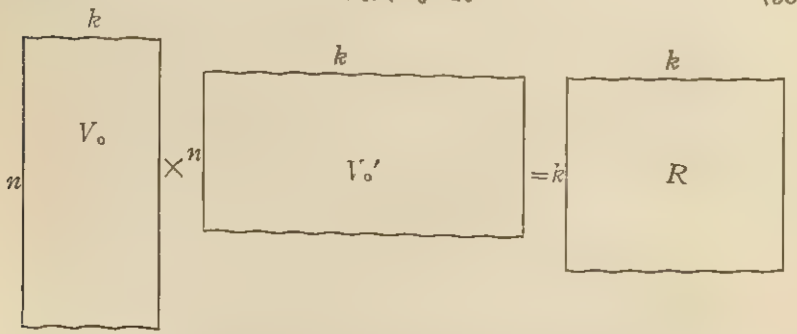
ations designed to approach a limit, as in principal components. Alternatively (and this is computationally the most efficient procedure) one proceeds through our last equations (52, 53) above, to

$$R(N) = S \cdot S' = P' \cdot P \tag{54}$$

factorizing R by conventional procedures. The virtue of direct factorization, however, remains that it can be used in situations for which R *cannot* be computed as above, namely, (a) when S is a three-way[8] score matrix and (b) when the scores in S are not given in numerical form. In the latter case, direct factorization (or K-way scale analysis, as Saunders calls it in this situation) avoids the Scylla of assuming normal distribution and the Charybdis of assuming that scores are in qualitatively different categories. With its minimum of assumptions it avoids the introduction of eccentricity (difficulty) factors and of other assumption factors. The reader interested in direct factor analysis or K-way scale analysis should read it at the source (**110**). At present it is practically untried, and its slightly greater freedom from assumptions does not justify, for general factor analysis, the extra labor involved. But it has value with nonquantitative (ranked) data, and in any case it reminds us of the fundamental matrix relation of S, V_{Fn}, and P.

The next most fundamental relation of which we need to be reminded is that the correlation matrix is obtained by postmultiplying the (test) factor matrix by its transpose, as follows:

$$V_0 \times V_0' = R \tag{55}$$



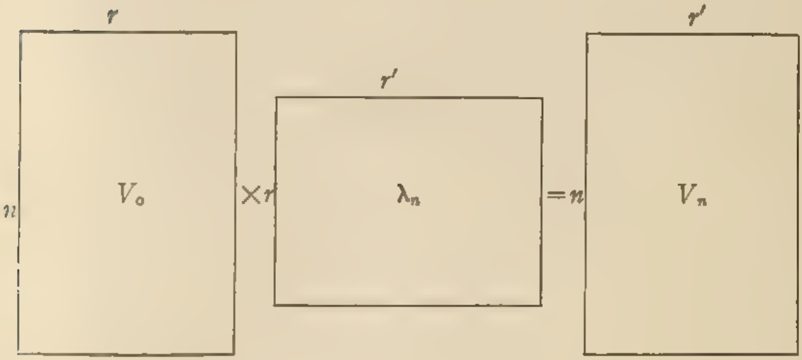
(where k is the number of factors)

[8] A three-way score matrix is one in which each score has three referents, e.g., a subject, a trait, and a judge. These may be indicated either by three subscripts to the number (score) or by arranging numbers in a three-dimensional matrix, as implied by the covariation chart.

This is our usual method of testing the goodness of V_0, by attempting to restore the correlations, and it cannot be done with the oblique factor matrix V_{Fn} without bringing in an extra matrix.

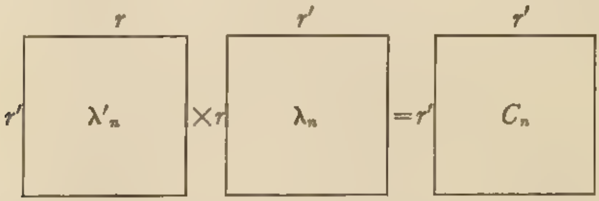
By now the student must be *thoroughly* familiar with the next matrix multiplication to be summarized: that describing the rotation process for k reference vectors when the transformation matrix λ shifts V_0 to simple structure, as follows: ($k=r$ will be used here to remind us that we deal with a reference vector matrix.)

$$V_0 \times \lambda_n = V_n \tag{56}$$



and

$$C_c = \lambda'_n \lambda_n \tag{57}$$

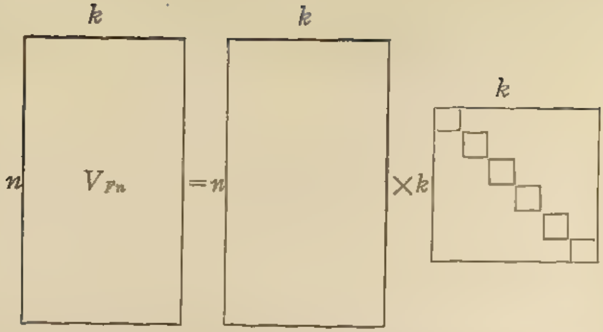


The latter gives us the angles among the reference vectors. To obtain the *factor* matrix V_{Fn} and the angles among the factors, C_{Fn}, we proceed as follows:

$$V_0 \times \lambda_{Fn} = V_{Fn} \tag{58}$$

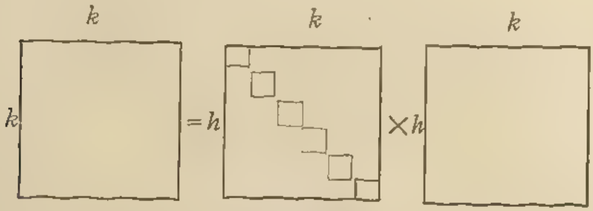
where

$$V_{Fn} = V_{Rn} D \tag{59}$$



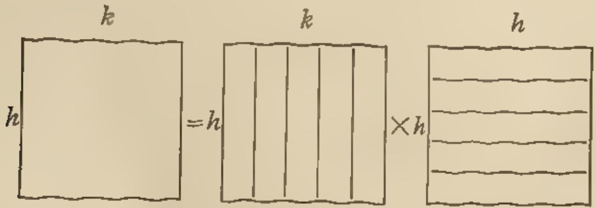
where D is a diagonal matrix which multiplies each column in V_{Rn} by the cosine of the angle between the factor, and the reference vector, D is obtained from the equation

$$\lambda_{Fn} = D\lambda_{Rn}^{-1} \tag{60}$$



which means that it is the series of values required to normalize the columns of the inverse of λ_{Rn} and can thus be obtained by calculating the latter from λ_{Rn}. Thus finally we obtain the angles among the factors

$$C_F = \lambda_F'\lambda_F = (D\lambda_R^{-1})'(D\lambda_R^{-1}) \tag{61}$$



The above equation summarizes all the key processes described in this book to obtain correlation matrices, factor loadings, rotated factor loadings and reference vector correlations, angles among factors and reference vectors, and the endowment of individuals in factors.

Questions and Exercises

1. State the four stages in a factor analysis (additional to the setting out of results) at which different time and labor-saving devices can be introduced. Indicate which can be most aided by special planning and which by straight machine aids.
2. When a large number of variables is involved indicate to what extent the time for (a) computation of correlations and (b) factor extraction can be reduced by each of three ways of breaking up the large correlation matrix.
3. Describe ways of shortening the computation of correlation coefficients and indicate the suitability of each to particular factorization enterprises.
4. Set out the principal steps in the calculation of (a) tetrachoric and (b) product moment correlations by I.B.M. equipment.
5. Describe two approaches to calculating the inverse of a triangular matrix.
6. Discuss the calculation of factor loadings (unrotated) by Dwyer's extension and Saunders' method and indicate briefly the principal steps in the calculation.
7. List, with brief description, all possible aids for shortening the process of rotation (a) when exploring a new field of data and (b) when making a factor resolution where previous research has reliably structured the field.
8. Set out six matrix multiplication equations (including a graphic indication of the arrangement of variables, etc. on the edge of each matrix) which summarize the principal transformations that are important to the factor analyst.

Glossary

Attenuation: deviation from theoretically true correlation due to experimental error.

Bifactor method of factorization: method of factorizing in which one looks for a general factor among all tests and a positive factor in each group.

Bimodal: having two modes (rather than one).

Bipolar factor: one having both positive and negative loadings, equally numerous.

Centroid: a center of gravity; an average point or position from which the sum of distances (with sign) of all observed points or positions is zero.

Centroid method of factorization: method of extracting factors in which the sum of all elements of each residual matrix is approximately zero before reflection (see Chapter 3).

Cluster: matrix, usually smaller than the original, representing tests or reflections of tests whose intercorrelations are high and positive.

Coefficient of pattern similarity: statistic used in matching two factor loading patterns (see page 306).

Common factor: statistical representation of some ability or trait which two or more items or tests in the battery have in common.

Common factor space: geometrical space of r dimensions, where r is the number of common factors obtained by analyzing the given data.

Common factor variance: synonym for communality.

Communality: sum of squares of factor loadings for any given item, i.e., total variance due to factors which this item shares with other items of the battery.

Configuration of points: arrangement representing relative positions of the test vectors in space by means of the coördinates of their end points. (See also Thurstone, **126,** Chapter 9 and page 91.)

Consistency coefficient: correlation between split halves (odd and even items) of a test all administered at the same time. Contrast reliability coefficient, in which an interval intervenes.

Constellation of points: general arrangement of the loadings among factors (Chapter 9).

Contingency table: one showing frequency of individuals classified according to two or more attributes about which we wish to observe possible correlations.

Coöperative factors: two, or more, factors which load similarly the same group of items, but in different proportions.

Correlation matrix: rectangular array of correlations, a_{ij}, where i and j represent the numbered positions of the correlated items as arranged by rows and columns, respectively. If the matrix is square, then for every pair (i, j), $a_{ij}=a_{ji}$.

Covariance: mean product of deviations of variable X and variable Y from their means: $(1/N)\ \Sigma(X-\overline{X})\ (Y-\overline{Y})$.

Covariance matrix: one whose elements are the covariations of the variables represented by its rows and columns.

Diagonal matrix: square matrix having zeros in all positions except those on the diagonal from upper left to lower right.

Direction cosine: one of a set of cosines of angles, defined for a point, each angle being measured between one of the reference axes and the vector connecting the point with the origin.

Discriminant function: a device which indicates how to combine a set of variables to give a total which will show the maximum difference or discriminative power between two groups.

Efficacy of a factor: number of situations in which a factor behaves as an indivisible entity.

Element of a matrix: single entry in a matrix.

Errors of measurement: errors arising mainly from the experimenter. These may include faulty observation, inaccurate interpretation of responses, giving ambiguous instructions and getting irrelevant answers, faulty recordings of responses, or errors on the part of the subject when he disregards instructions.

Factor configuration: see configuration and distinguish from factor structure.

Factor covariance matrix: matrix each of whose elements is the product of the factor loadings of the item in whose row and column it appears (see Chapters 3 and 4).

Factor fixation: Defining rotated factors by direction cosines in relation to a reference system. Rotation involves both "finding" and "fixing."

Factor invariance: (See Thurstone, Chapter 16.)

Factor loading: correlation of any particular test with the factor being extracted.

Factor matrix: matrix whose entries are the factor loadings obtained from a factor analysis; it generally is arranged so that it has as many columns as factors extracted and as many rows as tests in the original battery. Also referred to in Chapter 9 as a factor pattern matrix.

Factor resolution: the particular resolution into factors adopted for a given test configuration when the axes are rotated to some specific position.

Factor structure matrix: matrix of correlations between the variables and the factors; different from factor pattern if axes are oblique. (Chapter 9.) Used by Thurstone in the different sense labeled *factor resolution* here.

First-order factors: factors among single tests or variables.

Function fluctuation: excess of the error in the reliability coefficient over the consistency coefficient (see Chapter 6, page 85).

General factor: factor present in all tests of the battery.

Gramian matrix: matrix, such as a correlation matrix, in which the entry a_{ij} in the ith row, jth column is the same as that in the jth row, ith column (a_{ji}), and in which all principal minors have a value greater than or equal to zero.

Group method of factorization: one of the methods of factoring which uses only a portion of the variance of a matrix to determine the factor to be extracted (see Chapter 11).

Grouping method of factorization: (ibid.)

Hyperplane: space of $(n-1)$ dimensions, defined by a reference vector perpendicular to it; examples are: in two dimensions $(n=2)$ either coördinate axis is the hyperplane of the other; in three dimensions $(n=3)$ the plane defined by any two coördinate axes is the hyperplane of the third.

i, j: letters used to stand for any one of a known sequence of positive integers which are used in order values of a variable.

Identity matrix: diagonal matrix all of whose nonzero elements are equal to 1.

Interactive score units: Psychological or sociological scores that are not relative to other organisms scores but expressed directly in physical units describing the extent of interaction with the environment. This includes all "raw" scores.

Inverse of a matrix: matrix, M^{-1}, related to a given matrix M in such a way that the products $(M^{-1})(M)$ and $(M)(M^{-1})$ are both equal to the identity matrix.

Ipsative score units: units in which the raw scores have been expressed as standard scores with respect to a standard deviation of many scores within *an individual* instead of a population of persons.

Lambda (λ)*:* Greek letter corresponding to L.

Lambda (λ) *matrix:* name given in this book to the matrix of cosines which is used in rotating a factor matrix to a new position in the search for simple structure (see Chapter 12).

Maximise: make as large as possible.

Maximum likelihood, method of: A procedure for factor analysis developed by Lawley (**83**) and Young (**147**) which provides the best fitting factor matrix for a given number of factors. The degree of the goodness of the fit to the correlation matrix can then be rigorously evaluated in terms of a χ^2 statistic.

Method of coincidental markers: method of matching factors in two separate researches by comparing the number of markers used in one research which appear in a second research.

Minimize: make as small as possible.

Monotonic: a sequence of numbers is said to be monotonic if the numbers are arranged so that each is larger (or smaller) than the one preceding it in the sequence.

Multiple group method: (See Chapter 11.)

Multivariate selection: selecting a new set of data on the basis of homogeneity of more than one test.

Nonconstant errors: errors which do not affect everyone similarly.

Nonessential errors: errors arising from the guessing of communalities, use of correlations which have some measures missing, incompleteness of factor extraction, and the presence of computational errors.

Normalize: to divide each of a set of numbers by the square root of the *sum* of the squares of all numbers in the set, so that the sum of squares of the new set is 1.00.

Normative score units: Results of expressing raw scores (interactive units) in relation to those of the rest of the population (species) e.g. as percentiles, I.Q.'s etc.

Oblique: inclined at some angle other than 90°.

Order of matrix: description of the size of a matrix by the number of rows and columns it contains: that is, if it has m rows, n columns, it is then said to be of order m by n (written m by n).

Orthogonal: at right angles; perpendicular.

P-technique: (See Chapter 6.)

Parallel proportional profile: (See page 246.)

Personality sphere: concept in which all personality traits can be represented as though on a surface.

Phi coefficient: (See page 322.)

Plateau test: Used in two contexts: (1) Testing a meaningful rotation position by finding when a plateau of high loadings in factor contrasts with zero loadings. (2) Testing that rotation has reached its end point through the number in the hyperplane having failed to increase after three or four successive rotations i.e. having reached a plateau.

Primary factor: trait corresponding to the unit vector (*primary* vector) defined by a coördinate axis (see **126,** page 348.)

Principal axes: (See **126,** page 474).

Principle of orthogonal additions: (See page 247.)

Product matrix: result of multiplying together factor loadings of variables.

Projection of test vector: scalar product of test vector with the unit vector along the axis upon which the test vector is projected; (length of test vector × cos y, where y is the angle between the TV and the axis upon which it is being projected).

Q-technique: (See Chapter 6.)

R-technique: (See Chapter 6.)

Ramifying linkage method: method, used in group, grouping, or multiple-group methods of factoring, of obtaining clusters from a matrix of correlations.

Rank of matrix (related to number of factors)*:* number of rows (or columns) of the largest square matrix within the given matrix, whose determinant does not equal zero; the rank of the correlation matrix is theoretically equal to the number of common factors among the tests.

Reference vector: line of intersections of all hyperplanes except that with which we are concerned.

Reflect: change the signs of all matrix elements which indicate variance in the item to be reflected, i.e., the item such as sociability which is now to be regarded as unsociability.

Reliability coefficient of a test: correlation between results of two repetitions of the test, or between the results of two tests designed to measure the same traits.

Residual matrix: new matrix obtained by extracting variance due to a factor from a given matrix. (For discussion see Chapters 3 and 4.)

Rotation: process of moving factor axes and their hyperplanes in order to allow more points to fall in these hyperplanes.

Sampling error: tendency of any particular sample of population of persons (variables, etc.) to have a different mean and standard deviation in any measurement from the total or ideal population.

Second-order factors: factors among factors (among clusters of single tests or variables).

Sectional view rotation method: any rotation method in which graphing of (two-dimensional) cross sections of the factor space is used to help in finding the best hyperplane for each factor.

Simple factor: Term used by Holzinger and Harris to describe a Reference Vector, as contrasted with Primary Factor, by which they describe what we have called a Factor.

Simple structure: position of factor axes and their hyperplanes for which the maximum number of points possible has been rotated into each hyperplane.

Single-plane rotation method: method of rotation in which one hyperplane and its reference vector is fixed before any other reference vectors are shifted.

Specific factor: statistical representation of some ability or trait which only one item or test contains.

Specification equation: equation which indicates an individual's performance on a test in terms of loadings and factor endowments. (See Chapter 6.)

Standard deviation: σ; positive square root of variance.

Standard error of a loading: formula which indicates the amount of error in a factor loading (see page 293).

Submatrix method: Any method of factor extraction which re-arranges the order of the variables in the correlation matrix in order to bring clusters into distinct areas of the matrix and which then computes separately for these sub-matrices.

Test space: geometric space of $(n+r)$ dimensions, where r is the number of common factors; n is the number of specific factors of all the tests of the battery.

Test vector: vector from origin to a point in n-dimensional space.

Tetrad difference equation: Spearman's method of rearranging four correlation coefficients in combination with each other; this he considered as proof of the two-factor theory when the product was equal to or approximated zero.

Total centroid method of factorization: (See Chapter 10.)

Transformation matrix: matrix of cosines between reference vectors and factor axes, used in making a rotation.

Transpose (M′) of a matrix (M): matrix (M') whose successive rows are, in order, the successive columns of a given matrix (M).

Transposed factor analysis: Analysis of a correlation matrix by correlating rows instead of columns. Thus Q- is the transposed form of R-technique, O- of P-technique, and so on.

Triangular matrix: matrix containing all zeros in the portion above and to the right, or below and to the left, of the diagonal from upper left to lower right.

Unipolar factor: one having only positive or only negative loadings.

Unique factor: same as specific factor.

Unit covariance: the scaling of covariance from its observed value to 1.

Univariate selection: selecting a new set of data on the basis of their homogeneity with that of a single test in a previously factorized battery with the idea of factorizing the new data to compare with the old.

Unrotated matrix: name usually given to the matrix of factor loadings from which the first rotation toward simple structure is calculated.

V matrix: name given to any of the factor matrices used in a rotation to simple structure; when a subscript is applied (as V_3 matrix), this number tells how many rotations have already been performed. Thus, the V_0 matrix is the unrotated matrix, etc.

Variable: quantity which may take on several values in the course of a problem.

Variance: σ^2; sum of squares of deviations from their mean of a set of scores or other data.

Zero loadings: a factor loading so small that its magnitude is probably due entirely to chance or experimental error.

Appendix

ESSENTIAL STEPS IN MATRIX MULTIPLICATION BY ELECTRONIC DIGITAL CALCULATORS[1]

STATEMENT OF THE PROBLEM

Given a matrix V, of M rows and N columns, and a matrix λ of N rows and N columns, we wish to calculate the product matrix, $V' = V \times \lambda$.

GENERAL DISCUSSION

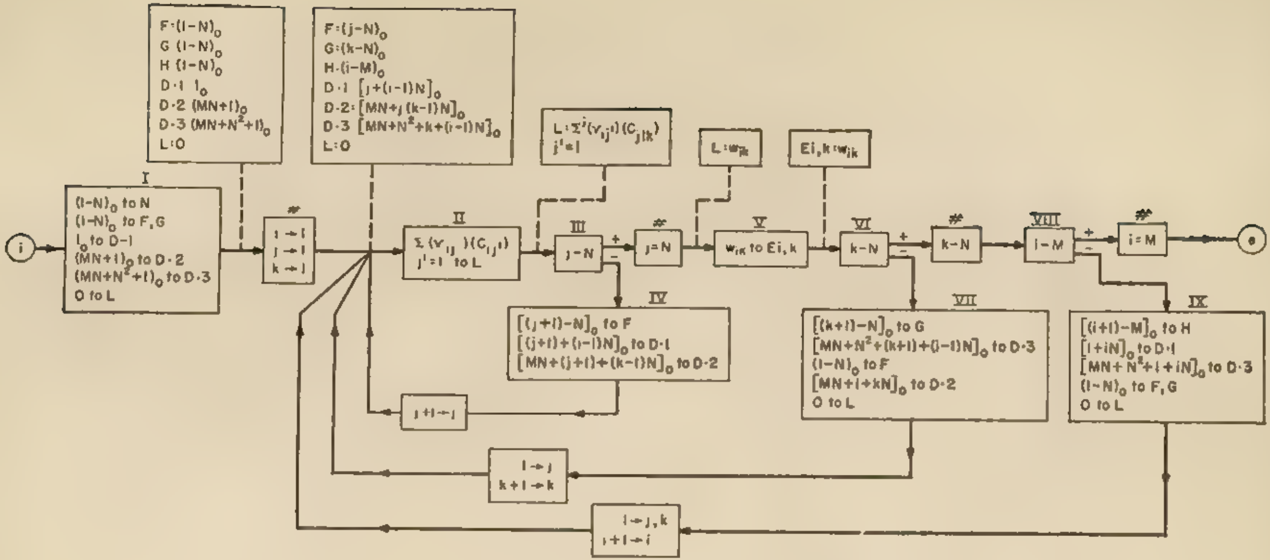
The size of the individual elements, v_{ij} and c_{jk}, will here present no machine problem since the range of both these sets of numbers is between -1 and $+1$. Also the elements, w_{ik}, of V', being projections of vectors within a unit hypersphere, upon axes having the center of this hypersphere as their origin, should still determine vectors within the same unit hypersphere, and will do so under the conditions of the rotations which we are performing. Thus the w_{ik} will also lie within the machine's range.

The size of the matrices, however, does present a problem in machine storage. In practice the value of N rarely exceeds 16 but the value of M may range all the way from about 20 to 80 or even 90, and hence in the cases involving many variables (large M) and also many factors (large N) it will be necessary to calculate V' in several steps. It has seemed best, since λ will always be comparatively small, to do any necessary subdividing on V rather than on λ. We shall then perform our multiplication by holding i (row subscript of V, V') constant until all the elements of the i^{th} row of V' have been calculated. (In the coding we shall use M as the number of rows of V which are being used in one round of calculations.) A discussion of the various possibilities arising from different sizes of V and λ will follow the flow diagram and static coding of the general case.

In the matrices to be considered, the individual elements of V and λ are independent of one another, but are known. We shall then need MN spaces to store the v_{ij} and N^2 spaces to store the c_{jk}, locating each space by means of its stored address. We shall also find it more convenient to use both our counting index and our addresses in the same machine form (i.e., that commonly used for addresses) since then we need store only a single form of 1, namely 1_o, which can then be used in stepping up either an address or an index.

[1] Contributed by Mildred Brannon, in charge of computing division one at the Laboratory of Personality Assessment and Group Behavior, University of Illinois.

Flow Diagram



Fixed Storage

A·i,j	:	v_{ij}
B·j,k	:	c_{jk}
C·1	:	1_0
C·2	:	$(1-M)_0$
C·3	:	$(1-N)_0$
C·4	:	$(MN+1)_0$
C·5	:	$(MN+N^2+1)_0$
C·6	:	O

Variable Storage

E·i,k	:	w_{ik}
D·1	:	$[j+(i-1)N]_0$
D·2	:	$[MN+j+(k+1)N]_0$
D·3	:	$[MN+N^2+k+(i-1)N]_0$
F	:	$(j-N)_0$
G	:	$(k-N)_0$
H	:	$(i-M)_0$
L	:	$\sum_{j'=1}^{j}(v_{ij'})(c_{j'k})$

The storage will be as follows:[2]

Fixed Storage

A.i,j : v_{ij}, where $\left\{\begin{matrix} i \leqslant M \\ j \leqslant N \end{matrix}\right\}$. Each A.$i,j$ will have $j+(i-1)N$ as its address.

B.j,k : c_{jk}, where $j,k \leqslant N$. Each B.j,k will have $MN+j+(k-1)N$ as its address.

C.1 : 1_o

C.2 : $(1-M)_o$

C.3 : $(1-N)_o$

C.4 : $(MN+1)_o$

C.5 : $(MN+N^2+1)_o$

C.6 : 0

Variable Storage

D.1 : $[j+(i-1)N]_o$

D.2 : $[MN+j+(k-1)N]_o$

D.3 : $[MN+N^2+k+(i-1)N]_o$

E.i,k : w_{ik}, where $\left\{\begin{matrix} i \leqslant M \\ k \leqslant N \end{matrix}\right\}$. Each E.$i,k$ will have $MN+N^2+k+(i-1)N$ as its address.

F : $(j-N)_o$

G : $(k-N)_o$

H : $(i-M)_o$

L : $\sum_{j'=1}^{j}(v_{ij'})(c_{j'k})$

CHOOSING M WITH RELATION TO N:

As mentioned earlier, we very rarely need to have the value of N greater than 16, but for some kinds of data we may need only an N as small as 4 or 5. In any case it is clear that as the value of N increases, the value for M must be decreased by a certain amount, thus allowing for sufficient space to store the calculated matrix. Since there are 1024 Memory storage spaces, and since we need 47 of them for coding and for storage of numbers other than the elements of the three matrices, we may relate M and N by the following equation:

$$2MN+N^2+47=1024$$

By algebra we then have: $$M=\frac{977-N^2}{2N},$$

from which we calculate the following table:

N	16	15	14	13	12	11	10	9	8	7	6	5
Maximum Integral M	22	25	27	31	34	38	43	49	57	66	78	95
Extra Memory Spaces	17	2	25	2	17	20	17	14	1	4	5	2
Possible extra w_{ik}	1	—	11	–	5	9	7	5	—	—	—	—

[2] Notation here used is in accordance with that of the Von Neumann and Goldstine reports **(134)**.

Static Coding

Order #	Order	Description of Order
I. 1	C.2	Accumulator contains $(i-M)_o$ - - - $(i=1)$
. 2	H S	H contains $(i-M)_o$
. 3	C.3	Accumulator contains $(j-N)_o$ - - - $(j=1)$
. 4	F S	F contains $(j-N)_o$
. 5	G S	G contains $(k-N)_o$ - - - $(k=1)$
. 6	C.1	Accumulator contains 1_o
. 7	D.1 S	D.1 contains 1_o - - $([j+(i-1)N]_o=1)$
. 8	C.4	Accumulator contains $(MN+1)_o$ - - $[j+(k-1)N=1]$
. 9	D.2 S	D.2 contains $(MN+j+(k-1)N)_o$
.10	C.5	Accumulator contains $(MN+N^2+1)_o$ - - $[k+(i-1)N_o=1]$
.11	D.3 S	D.3 contains $(MN+N^2+k+(i-1)N)_o$
.12	C.6	Accumulator contains 0
.13	L S	L contains 0
II. 1	D.1	Accumulator contains $(j+(i-1)N)_o$
. 2	II.3 Sp'	(Address in II.3 will be that of v_{ij})
. 3	(A.i,j)R	Register contains v_{ij}
. 4	D.2	Accumulator contains $(MN+j+(k-1)N)_o$
. 5	II.6 Sp	(Address in II.6 will be that of c_{jk})
. 6	$(B.j,_k)X$	Accumulator contains first 39 digits of $(v_{ij})(c_{jk})$
. 7	L $h+$	Accumulator contains $\sum_{j'=1}^{j}(v_{ij'})(c_{j'k})$
. 8	L S	L contains $\sum_{j'=1}^{j}(v_{ij'})(c_{j'k})$

Order #	Order		Description of Order
III. 1	F		Accumulator contains $(j-N)$.
.2	V.1	Cc	Controls to (left-hand) order V.1 if $j=N$
IV. 1	C.1	$h+$	Accumulator contains $(j+1-N)$.
.2	F	S	F contains $(j+1-N)$.
.3	D.1		Accumulator contains $(j+(i-1)N)$.
.4	C.1	$h+$	Accumulator contains $(j+1+(i-1)N)$.
.5	D.1	S	D.1 contains $(j+1+(i-1)N)$.
.6	D.2		Accumulator contains $(MN+j+(k-1)N)$.
.7	C.1	$h+$	Accumulator contains $(MN+j+1+(k-1)N)$.
.8	D.2	S	D.2 contains $(MN+j+1+(k-1)N)$.
.9	II.1	C'	Controls to order II.1
V. 1	D.3		Accumulator contains $(MN+N^2+k+(i-1)N)$.
.2	V.4	Sp'	(Address at V.4 will be that of w_{ik})
.3	L		Accumulator contains w_{ik}
.4	(E.i,k)S		E.i,k contains w_{ik}
.5	II.1	C'	Controls to order II.1
VI. 1	G		Accumulator contains $(k-N)$.
.2	VIII.1	Cc'	Controls to (right-hand) order VIII.1 if $k=N$
VII. 1	C.1	$h+$	Accumulator contains $(k+1-N)$.
.2	G		G contains $(k+1-N)$.
.3	D.3		Accumulator contains $(MN+N^2+k+(i-1)N)$.
.4	C.1	$h+$	Accumulator contains $(MN+N^2+k+1+(i-1)N)$.
.5	D.3	S	D.3 contains $(MN+N^2+k+1+(i-1)N)$.

Order #	Order		Description of Order
.6	C.3		Accumulator contains $(1-N)$.
.7	F	S	F contains $(1-N)$.
.8	C.6		Accumulator contains 0
.9	L	S	L contains 0
.10	D.2		Accumulator contains $(MN+j+(k-1)N)$.
.11	C.1	$h+$	Accumulator contains $(MN+1+kN)$.
.12	D.2	S	D.2 contains $(MN+1+kN)$.
VIII. 1	H		Accumulator contains $(i-M)$.
.2	e	Cc'	End, if $i=M, j=k=N$
IX. 1	C.1	$h+$	Accumulator contains $(i+1-M)$.
.2	H	S	H contains $(i+1-M)$.
.3	D.1		Accumulator contains $(j+(i-1)N)$.
.4	C.1	$h+$	Accumulator contains $(1+iN)$.
.5	D.1	S	D.1 contains $(1+iN)$.
.6	D.3		Accumulator contains $(MN+N^2+k+(i-1)N)$.
.7	C.1	$h+$	Accumulator contains $(MN+N^2+1+iN)$.
.8	D.3	S	D.3 contains $(MN+N^2+1+iN)$.
.9	C.3		Accumulator contains $(1-N)$.
.10	F	S	F contains $(1-N)$.
.11	G	S	G contains $(1-N)$.
.12	C.6		Accumulator contains 0
.13	L	S	L contains 0
.14	II.1	C'	Controls to order II.1

Coding Sequence

Memory Space #	Contents
1	1_o
2	$(1-M)_o$
3	$(1-N)_o$
4	$(MN+1)_o$
5	$(MN+N^2+1)_o$
6	0
7	$(j+(i-1)N)_o$
8	$(MN+j+(k-1)N)_o$
9	$(MN+N^2+k+(i-1)N)_o$
10	$(j-N)_o$
11	$(k-N)_o$
12	$(i-M)_o$
13	$\sum_{j'=1}^{j}(v_{ij'})(c_{j'k})$
14	(3. . . . . ,12 S . . . .)
15	(2. . . . . ,10 S . . . .)
16	(11 S . . . . ,1 . . . .)
17	(7 S . . . . ,4 . . . .)
18	(8 S . . . . ,5 . . . .)
19	(9 S . . . . ,6 . . . .)
20	(13 S . . . . ,7 . . . .)
21	(21 Sp' . . . . ,$(47+j+(i-1)N)$ R)
22	(8 . . . . ,23 Sp . . . .)
23	($(47+MN+j+(k-1)N)$ X . . . . . . ,13 $h+$ . . . .)
24	(13 S . . . . ,7 . . . .)
25	(30 Cc . . . . ,1 $h+$ . . . .)
26	(10 S . . . . ,7 . . . .)
27	(1 $h+$ . . . . ,7 S . . . .)
28	(8. . . . . ,1 $h+$ . . . .)
29	(8 S . . . . ,20 C' . . . .)
30	(9. . . . . ,31 Sp'. . . . . . .)
31	(13 . . . . ,$(47+MN+N^2+k+(i-1)N)$ S) . . .)
32	(20 C' . . . . ,11 . . . .)
33	(39 Cc' . . . . ,1 $h+$ . . . .)
34	(11 S . . . . ,9 . . . .)
35	(1 $h+$ . . . . ,9 S . . . .)
36	(2. . . . . ,10 S . . . .)
37	(6. . . . . ,13 S . . . .)
38	(8. . . . . ,1 $h+$ . . . .)
39	(8 S . . . . ,12 . . . .)
40	(47 Cc' . . . . ,1 $h+$ . . . .)
41	(12 S . . . . ,7 . . . .)
42	(1 $h+$ . . . . ,7 S . . . .)
43	(9. . . . . ,1 $h+$ . . . .)
44	(9 S . . . . ,3 . . . .)
45	(10 S . . . . ,11 S . . . .)
46	(6. . . . . ,13 S . . . .)
47	(20 C' . . . . ,e . . . .)

48 - - $(47+MN)$: $A.i,j$
$(47+MN+1)$ - - $(47+MN+N^2)$: $B.j,k$
$(47+MN+N^2+1)-(47+2MN+N^2)$: $E.i,k$

NOTE: Although the last three groups of addresses involve large numbers, we shall choose M and N, by a table to be described, so that no address will exceed $1024=2^{10}$ in numerical value.

Since the relation between M and N is not linear but is quadratic, we find the values of M increasing more rapidly with small values of N than with large values. We note too that for six values of N, there appears to be some space left in which may be stored an extra row of the V matrix and a few elements of the $(i+1)^{st}$ row of V'. If this were possible, it might conceivably save a third round of calculations, as for instance, if $M=55$, $N=14$, the calculation of 7 extra w_{ik} on each of two rounds of multiplication with $M=27$ each time, would then complete the problem without a third set of multiplications. However, such a scheme would necessitate several extra orders, such as one to end the calculation on the $(i+1)^{st}$ row of V with the element of that row whose column position is $N/2$, and one would need to allow more Memory space for such extra orders as well as for the extra calculated numbers. On the whole, therefore, it would seem inadvisable to calculate incomplete rows of V' unless many products were to be formed using the same set of directions for this extra set of values.

NOTE ON TIME FACTOR FOR ONE COMPLETE MULTIPLICATION:

Using the estimates of time for the various machine operations as given in the Von-Neumann-Goldstine report, we have the following table:

Box #	Total Operation Time	Number of Iterations
I	$(13\times25)\mu=325\ \mu$	1
II	$(25\times6+100+30)\mu=280\ \mu$	N^2M
III	$(25+30)\mu=55\mu$	N^2M
IV	$(25\times6+30\times3)\mu=240\ \mu$	N^2M
V	$(25\times5)\mu=125\ \mu$	NM
VI	$(25+30)\mu=55\ \mu$	NM
VII	$([25\times9]+[30\times3])\mu=315\ \mu$	NM
VIII	$(25+30)\mu=55\ \mu$	M
IX	$(25\times11+30\times3)\mu=365\ \mu$	M

In addition, we have 34 words each containing two orders, for which the time will be approximately $(34\times20)\mu=680\mu$.

We then have the total time expressed in this sum:

$$[680+325+(280+55+240)N^2M+(125+55+315)NM+(55+365)M]\ \mu$$
$$=[1005+5M(115N^2+99N+84)]\ \mu \quad .$$

We may assume that the maximum value of M will not exceed 100, and that that of N will not exceed 20. Using these figures, our sum becomes:

$$[1005+500(46000+1980+84)]\mu=24{,}033{,}005\ \mu \quad .$$

Since $\mu=(1\text{ sec})10^{-6}$, we can interpret this result to mean that the maximum time for a complete matrix multiplication of the type discussed here will be about 24 seconds, according to the estimates of time now available.

CONCLUSION

Although perhaps too much storage space is necessary for known values of the elements of V and of λ, to make this matrix multiplication an especially economical problem for the machine, yet if several such multiplications were to be performed, and if the matrices were fairly large, this would seem to be a much faster method of computing the product matrix than the use of the desk computers or other similar equipment.

Bibliography

1. Andree, R. V., A computational short cut in factor analysis, *Psychol. Bull.*, 1952, **49.**

1a. Aristotle, Metaphysics, in *The Basic Works of Aristotle* (R. McKeon, ed.), New York, Random House, 1941.

2. Baron, M., and Harper, R., The relation between the mechanical properties of Chesire cheese and graders' judgments, *Dairy Ind.,* 1950, **4,** 407-410.

3. Bartlett, M. S., The statistical conception of mental factors, *Brit. J. Psychol.,* (Statistical Section) 1948, **1,** 204-218.

4. Bechtoldt, D. C., An empirical study of the stability of a simple structure multiple factor solution, *Am. Psychol.,* 1949, **4,** 352.

5. Bechtoldt, D. C., The use of geometrical concepts of multiple factor analysis in multiple correlation problems, *Am. Psychol.,* 1949, **4,** 246.

6. Borden, E. S., Factor analysis in experimental designs in clinical and social psychology, *Psychol. Rev.,* 1943, **50,** 415-429.

7. Brogden, H. E., Variation in test validity with variation in the distribution of item difficulties, number of items, and degree of their correlation, *Psychometrika,* 1946, **11,** 197-214.

8. Brunswik, E., Systematic and representative design of psychological experiments with results in physical and social perception, *University of California Syllabus Series 304,* Berkeley, University of California Press, 1947.

9. Burt, C. L., Correlations between persons, *Brit. J. Psychol.,* 1937, **28,** 59-96.

10. Burt, C. L., Factor analysis by sub-matrices, *J. Psychol.,* 1938, **6,** 339-375.

11. Burt, C. L., *Factors of the Mind,* London, University of London Press, 1940.

12. Burt, C. L., A comparison of factor analysis and analysis of variance, *Brit. J. Psychol.,* (Statist. Sec.) 1948, **1,** 3-26.

13. Burt, C. L., Subdivided factors, *Brit. J. Psychol.,* (Statist. Sec.) 1949. **2,** 41-63.

14. Burt, C. L., Alternative methods of factor analysis and their relations to Pearson's method of principal axes, *Brit. J. Psychol.,* (Statist. Sec.) 1949, **2,** 98-122.

14a. Burt, C. L., A comparison of factor analysis and analysis of variance. *Brit. J. Psychol.,* (Statis. Sec.) 1947, **1,** 3-27.

14b. Carroll, J. B., The factor analysis of tests which differ in difficulty and whose scores are affected by chance success, in press.

15. Castone, G. F., and Dye, W. S., A simplified punch card method of determining sums of squares and sums of products, *Psychometrika,* 1949, **14,** 243-251.

16. Cattell, R. B., A note on correlation clusters and cluster search methods, *Psychometrika,* 1944, **9,** 169-184.

17. Cattell, R. B., Psychological measurement: normative, ipsative, interactive, *Psychol. Rev.*, 1944, **51**, 292-303.
18. Cattell, R. B., Parallel proportional profiles and other principles for determining the choice of factors by rotation, *Psychometrika*, 1944, **9**, 267-283.
19. Cattell, R. B., The principal trait clusters for describing personality, *Psychol. Bull.*, 1945, **42**, 129-161.
20. Cattell, R. B., The cultural functions of social stratification: I: regarding the genetic bases of society, *J. Soc. Psychol.*, 1945, **21**, 3-23.
21. Cattell, R. B., Simple structure in relation to some alternative factorizations of the personality sphere, *J. Genet. Psychol.*, 1946, **35**, 225-238.
22. Cattell, R. B., *The Description and Measurement of Personality*, New York, World Book, 1946.
23. Cattell, R. B., with Cattell, A. K. S., and Rhymer, R. M., P-technique demonstrated in determining psycho-physiological source traits in normal individuals, *Psychometrika*, 1947, **12**, 267-288.
24. Cattell, R. B., The integration of factor analysis with psychology, *J. Educ. Psychol.*, 1948, **39**, 227-236.
25. Cattell, R. B., Concepts and methods in the measurement of group syntality, *Psychol. Rev.*, 1948, **55**, 48-63.
26. Cattell, R. B., r_p and other coefficients of pattern similarity, *Psychometrika*, 1949, **14**, 279-298.
27. Cattell, R. B., The dimensions of culture patterns by factorization of national characters, *J. Abn. Soc. Psychol.*, 1949, **44**, 443-469.
28. Cattell, R. B., with Tiner, L. G., The varieties of structural rigidity, *J. Person.*, 1949, **17**, 321-341.
29. Cattell, R. B., A note on factor invariance and the identification of factors, *Brit. J. Psychol.* (Statis. Sec.), 1949, 134-139.
30. Cattell, R. B., *Personality: A Systematic Theoretical and Factual Study*, New York, McGraw-Hill, 1950.
31. Cattell, R. B., and Saunders, D. R., The matching of personality factors from behavior rating, questionnaire, and test data, *J. Soc. Psychol.*, 1950, **31**, 243-260.
32. Cattell, R. B., and Adelson, M., The dimensions of social change in the U.S.A. as determined by P-technique, *Soc. Forces*, 1951.
33. Cattell, R. B., and Winder, A. P., Structural rigidity in relation to learning theory and clinical psychology, *Psychol. Rev.*, 1952.
33a. Cattell, R. B., The three basic factor analytical designs—P-, R-, and T-techniques—and their derivatives, *Psychol. Bull.*, 1952, **49**.
34. Cattell, R. B., Cattell, A. K. S., and Miller, A., An analytical solution and a practical test of rotation for parallel proportional profiles, 1952.
34a. Cattell, R. B., and Cross P., The use of P-technique for revealing individual dynamic and self structure, 1952.

35. Chesire, L., Saffir, M., and Thurstone, L. L., *Computing Diagrams for the Tetrachoric Correlation Coefficient,* Chicago, University of Chicago Bookstore, 1933.
36. Clendening, L., *The Human Body,* New York, Knopf, 1927, p. 150 ff.
37. Cohen, Jozef, Color vision and factor analysis, *Psychol. Rev.,* 1949, **56,** 224-233.
38. Cohen, M. R., and Nagel, E., *An Introduction to Logic and Scientific Method,* New York, Harcourt, Brace, 1934.
39. Coombs, C. H., A criterion for the significance of common factor variance, *Psychometrika,* 1941, **6,** 267-272.
40. Crout, P. D., A short method for evaluating determinants and solving systems of linear equations with real or complex coefficients, *Trans. Am. Inst. Elec. Engrs.,* 1941, **60,** 1235-1240.
41. Cureton, E. E., and Dunlap, J. W., Some effects of heterogeneity on the theory of factors, *Am. J. Psychol.,* 1930, **42,** 608-620.
42. Cureton, T. K., *Physical Fitness Appraisal and Guidance* (Appendix), St. Louis, Mosby, 1947.
43. Davies, M., The general factor in correlations between persons, *Brit. J. Psychol.,* 1939, **29,** 404-421.
44. Dwyer, P. S., The determination of the factor loadings of a given test from the known factor loadings of other tests, *Psychometrika,* 1937, **2,** 173-178.
45. Dwyer, P. S., The contribution of an orthogonal multiple factor solution to multiple correlation, *Psychometrika,* 1939, **4,** 163-171.
46. Dwyer, P. S., The evaluation of multiple and partial correlation coefficients from the factorial matrix, *Psychometrika,* 1940, **5,** 211-232.
46a. Dwyer, P. S., *Linear Computations,* New York, Wiley, 1951.
47. Emmett, W. G., Sampling error and the two factor theory, *Brit. J. Psychol.,* 1936, **26,** 362-387.
48. Emmett, W. G., Factor analysis by Lawley's method of maximum likelihood, *Brit. J. Psychol.* (Statis. Sec.), 1949, **2,** 90-97.
49. Eysenck, H. J., Criterion analysis: an application of the hypothetico-deductive method to factor analysis, *Psychol. Rev.,* 1950, **57,** 38-65.
49a. Eysenck, H. J., Critical notice of "primary mental abilities," *Brit. J. Educ. Psychol.,* 1939, **9,** 270-275.
50. Feigl, H., and Sellers, W., *Readings in Philosophical Analysis,* New York, Appleton, Century, Crofts, 1949.
51. Ferguson, G. A., The factorial interpretation of test difficulty, *Psychometrika,* 1941, **6,** 323-329.
52. Fisher, R. A., *The Design of Experiments,* Edinburgh, Oliver and Boyd, 1937.
53. Fiske, D. W., Consistency of the factorial structure of personality ratings from different sources, *Am. Psychol.,* 1948, **3,** 360.
54. Frazer, R. A., Duncan, W. J., and Collar, A. R., *Elementary Matrices and Some Applications to Dynamics and Differential Equations,* Cambridge, The University Press, 1938.

55. Fruchter, B., Effect of various weightings of wrong responses on the factorial content of speeded tests. *Am. Psychol.*, 1949, **4**, 306.
56. Fruchter, B., Note on the computation of the inverse of a triangular matrix, *Psychometrika*, 1949, **14**, 89-94.
57. Garnett, J. C. M., On certain independent factors in mental measurement, *Proc. Roy. Soc. London*, 1919, **96**, 91-111.
58. Guilford, J. P., The difficulty of a test and its factor composition, *Psychometrika*, 1941, **6**, 67-77.
59. Guilford, J. P., with Zimmerman, W. S., Some A. A. F. findings concerning aptitude factors, *Occupations*, 1947, **26**, 154-159.
60. Guttman, L., General theory and methods for matric factoring, *Psychometrika*, 1944, **9**, 1-16.
61. Hall, D. M., Welker, E. L., and Crawford, I., Factor analysis calculations by tabulating machines, *Psychometrika*, 1945, **10**, 93-125.
62. Hamilton, M., Nomogram for the tetrachoric correlation coefficient, *Psychometrika*, 1948, **13**, 259-271.
63. Hammond, W. H., Factor analysis as applied to social and economic data, *Brit. J. Educ. Psychol.*, 1946, **16**, 178.
64. Hargreaves, H. L., The faculty of imagination, *Brit. J. Psychol. Monogr. Suppl., No. 10*, 1927.
65. Harman, H. H., On the rectilinear prediction of oblique factors, *Psychometrika*, 1941, **6**, 29-35.
66. Harper, R. et al., The application of multiple factor analysis to industrial test data, *Brit. J. Applied Phys.*, 1950, **1**, 1-6.
67. Harris, C. W., Direct rotation to primary structure, *J. Educ. Psychol.*, 1948, **39**, 449-468.
68. Harris, C. W., with Knoell, D. M., The oblique solution in factor analysis, *J. Educ. Psychol.*, 1948, **39**, 385-405.
69. Harsh, C. M., Constancy and variation in patterns of factor loadings, *J. Educ. Psychol.*, 1940, **31**, 335-359.
70. Herdan, G., The logical and analytical relationship between the theory of accidents and factor analysis, *J. Roy. Statist. Soc.*, 1943, **106**, 125-142.
71. Holzinger, K. J., and Harman, H. H., *Factor Analysis*, Chicago, University of Chicago Press, 1941.
72. Holzinger, K. J., and Harman, H. H., A comparison of the principal axis and centroid factor, *J. Educ. Psychol.*, 1946, **37**, 449-472.
73. Holzinger, K. J., and Harman, H. H., Factoring factors, *J. Educ. Psychol.*, 1947, **38**, 321-328.
74. Holzinger, K. J., and Harman, H. H., Applications of the simple method of factor analysis, *J. Educ. Psychol.*, 1949, 40, 129-140.
75. Horst, P., *The Prediction of Personal Adjustment*, Bull. Soc. Sci. Res. Council, 1941, **48**.
76. Hotelling, H., Simplified calculation of principal components, *Psychometrika*, 1935, **1**, 27-35.
77. Hsü, E. H., An experimental demonstration of factor analysis, *J. Genet. Psychol.*, 1948, **38**, 235-241.

78. Hsü, E. H., The intrapersonal factor and its clinical applicability, *J. Person.*, 1949, **17,** 273-287.

79. Jackson, D., The geometry of frequency functions, *Bull. Am. Math. Soc.*, 1925, **31,** 63-73.

80. Kelley, T. L., *Essential Traits of Mental Life*, Cambridge, Harvard University Press, 1935.

81. Kendall, M. G., A new measure of rank correlation, *Biometrika*, 1938, **30,** 81-93.

82. Klüver, H., Psychology at the beginning of World War II, *J. Psychol.*, 1949, **28,** 383-410.

82a. Landahl, H. D., Centroid orthogonal transformations, *Psychometrika*, 1938, **3,** 219-223.

83. Lawley, D. N., The application of the maximum likelihood method to factor analysis, *Brit. J. Psychol.*, 1943, **33,** 172-175.

84. Lawley, D. N., The factorial analysis of multiple item tests, *Proc. Roy. Soc. Edinburgh*, 1944, **62,** 74-82.

85. Lazarsfeld, P. I., Factor analysis of qualitative attributes, *Am. Psychol.*, 1942, **2,** 306.

86. Loevinger, J., A systematic approach to the construction and evaluation of tests of ability, *Psychol. Monogr.*, 1947, **61.**

87. London, I. D., The role of the model in explanation, *J. Genet. Psychol.*, 1949, **74,** 165-176.

88. London, I. D., The concept of the behavioral spectrum, *J. Genet. Psychol.*, 1949, **74,** 177-184.

89. Lorge, I., Computational technics, *Rev. Educ. Res.*, 1945, **15,** 441-446.

90. Lovell, C., The effect of special construction of test items on their factor composition, *Psychol. Monogr.*, 1944, **56,** 6, 1-26.

91. Luborsky, L. B., and Hornadya, J., A mechanical factor-rotator for demonstration, *Am. J. Psychol.*, 1948, **61,** 104-106.

92. McCloy, C. H., Metheny, E., and Knott, V., A comparison of the Thurstone method of multiple factors with the Hotelling method of principal components, *Psychometrika*, 1938, **3,** 61-67.

93. McMahon, J., Hyperspherical geometry and its application to correlation theory for *n* variables, *Biometrika*, 1923, **15,** 173-208.

94. McNemar, Q., On the number of factors, *Psychometrika*, 1942, **7,** 9-18.

95. McNemar, Q., *Psychological Statistics*, New York; Wiley, 1949.

95a. McQuitty, L. L., A measure of individual differences in personality integration, *Canad. J. Psychol.*, 1950, **18,** 461-482.

96. Michaels, W. B., Factor analysis of tests and criteria: a comparative study of two A. A. F. pilot populations, *Psychol. Monogr. 298*, **63,** 1949.

97. Mosier, C. I., Influence of chance error on simple structure: an empirical investigation of the effect of chance error and estimated communalities on simple structure in factorial analysis, *Psychometrika*, 1939, **4,** 33-44.

97a. Mosier, C. I., and Uhlaner, J. E., Application of factorial logic to spatial abilities, A.P.A. Meeting, 1951 (In press).

98. Neumann, J. von, and Morgenstern, O., *The Theory of Games and Economic Behavior*, Princeton, Princeton University Press, 1947.

99. Noüy, Lecomte du, *Human Destiny*, New York, Longmans, 1947.

100. Peters, C. C., Multiple factor analysis, in *Encyclopedia of Psychology* (P. L. Harriman, ed.), New York, Philosophical Library, 1949.

100a. Peters, C. C., and Van Voorhis, W. R., *Statistical Procedures and Their Mathematical Bases*, New York, McGraw-Hill, 1940.

101. Primoff, E. S., Correlations and factor analysis of the abilities of the single individual, *J. Genet. Psychol.*, 1943, **28**, 121-132.

102. Reyburn, H. A., and Taylor, J. G., Some factors of personality, *Brit. J. Psychol.*, 1939, **30**, 151-165.

103. Rimoldi, H. J. A., The central intellective factor (In press, *J. Abn. Soc. Psychol.*).

104. Roff, M., The relation between results obtainable with raw and corrected correlation coefficients in multiple factor analysis, *Psychometrika*, 1937, **2**, 35-39.

105. Rosner, B., An algebraic solution for the communalities, *Psychometrika*, 1948, **13**, 181-184.

106. Sandler, J., The reciprocity principle as an aid to factor analysis, *Brit. J. Psychol.* (Statist. Sec.), 1949, **2**, 180-187.

107. Saunders, D. R., Factor analysis I. Some effects of chance error, *Psychometrika*, 1948, **13**, 251-257.

108. Saunders, D. R., Factor analysis II. A note concerning rotation of axes to simple structure, *Educ. Psychol. Meas.*, 1949, **9**.

109. Saunders, D. R., Factor analysis III. Desiderata in experimental design (In press).

110. Saunders, D. R., Practical methods in the direct factor analysis of psychological score matrices, *Ph.D. Thesis*, 1950, University of Illinois Library.

111. Smith, R. G., A factorial study of attitude toward the Negro, *Ph.D. Thesis*, 1950, University of Illinois Library.

112. Spearman, C., General intelligence objectively determined and measured, *Am. J. Psychol.*, 1904, **15**, 201-293.

113. Spearman, C., *The Abilities of Man: The Nature and Measurement*, London, Macmillan, 1927.

114. Spearman, C., Material versus abstract factors in correlation, *Brit. J. Psychol.*, 1927, **17**, 322-326.

115. Spearman, C., Abilities as sums of factors, or as their products, *J. Educ. Psychol.*, 1937, **28**, 629-631.

116. Stephenson, W., Tetrad differences for verbal sub-tests relative to nonverbal sub-tests, J. Educ. Psychol., 1931, **22**, 334-350.

117. Stephenson, W., The inverted factor technique, *Brit. J. Psychol.*, 1936, **26**, 344-361.

118. Swineford, F., Comparisons of the multiple-factor and bifactor methods of analysis, *Psychometrika*, 1941, **6**, 375-382.

119. Thomson, G. H., A hierarchy without a general factor, *Brit. J. Psychol.*, 1916, **8**, 271-281.
120. Thomson, G. H., *The Factorial Analysis of Human Ability*, London, University of London Press, 2d ed., 1946, p. 168.
121. Thomson, G. H., On estimating oblique factors, *Brit. J. Psychol.* (Statis. Sec.) 1949, **2**, 1-2.
122. Thorndike, E. L., *Your City*, New York, Harcourt Brace, 1939.
123. Thorndike, R. L., Individual differences, *Annual Review of Psychology*, Vol. I, 1950, 87-104.
124. Thouless, R. H., The factorial analysis of human abilities: a reply, *The Human Factor*, 1935, **9**, 180-185.
125. Thurstone, L. L., *Primary Mental Abilities*, Chicago, Chicago University Press, 1938.
126. Thurstone, L. L., *Multiple Factor Analysis*, Chicago, Chicago University Press, 1947.
127. Travers, R. M. W., The use of a discriminant function in the treatment of psychological group differences, *Psychometrika*, 1939, **4**, 25-32.
128. Tryon, R. C., *Cluster Analysis*, Berkeley, University of California Press, 1939.
129. Tucker, L. R., A matrix multiplier, *Psychometrika*, 1940, **5**, 289-294.
130. Tucker, L. R., A semi-analytical method of factorial rotation to simple structure, *Psychometrika*, 1944, **9**, 43-48.
131. Tucker, L. R., A note on the computation of a table of intercorrelations, *Psychometrika*, 1948, **13**, 245-249.
132. Tucker, L. R., Simplified punched card methods in factor analysis, *Research Forum*, Aug. 26, 1946, New York, I. B. M. Corp.
133. Vernon, P. E., and Parry, J. B., *Personnel Selection in the British Forces*, London, University of London Press, 1949.
134. Von Neumann, J., and Goldstine, H. H., Report on the mathematical and logical aspects of an electronic computing instrument, Princeton, N. J., Institute for Advanced Study, 1947.
135. Waugh, F. V., and Dwyer, P. S., Compact computation on the inverse of a matrix, *Am. Math. Statist.*, 1945, **16**, 259-271.
136. Webb, E., Character and intelligence, *Brit. J. Psychol., Monogr. Suppl. No. 3*, London, 1915.
137. Wherry, R. J., and Gaylord, R. H., Factor pattern of test item and tests as a function of the correlation coefficient: content, difficulty, and constant error factors, *Psychometrika*, 1944, **9**, 237-244.
138. Wherry, R. J., A new iterative method for correcting erroneous communality estimates in factor analysis, *Psychometrika*, 1947, **14**, 231-243.
139. Whitfield, J. W., Rank correlation between two variables, one of which is ranked, the other dichotomous, *Biometrika*, 1947, **34**, 292-296.
140. Weiner, N., *Cybernetics or Control and Communication in the Animal and the Machine*, Cambridge, Mass., Technology Press, 1948.

141. Williams, H. V. M., A P-technique study of personality factors in the psychosomatic area, *Ph. D. Thesis,* 1949, University of Illinois Library.

141a. Winch, R. F., Some data bearing on the Oedipus hypothesis, *J. Abn. and Soc. Psychol.,* 1950, **45,** 481-489.

142. Wittenborn, J. R., Factorial equations for tests of attention, *Psychometrika,* 1943, **8,** 19-35.

143. Woodrow, H., Interrelationships of measures of learning, *J. Psychol.,* 1940, **10,** 49-73.

144. Woodrow, H., and Wilson, L. A., A simple procedure for approximate factor analysis, *Psychometrika,* 1936, **1,** 245-258.

145. Young, G., Factor analysis and the index of clustering, *Psychometrika,* 1939, **4,** 201-208.

146. Young, G., Maximum likelihood estimation and factor analysis, *Psychometrika,* 1941, **6,** 49-53.

147. Young, G., with Householder, A. S., Factorial invariance and significance, *Psychometrika,* 1940, **5,** 47-56.

148. Zimmerman, W. S., A simple graphical method for orthogonal rotation of axes, *Psychometrika,* 1946, **11,** 51-55.

Indexes

INDEX OF NAMES

INDEX OF SUBJECTS